(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization International Bureau





(43) International Publication Date 16 May 2002 (16.05.2002)

PCT

(10) International Publication Number WO 02/38613 A2

- (51) International Patent Classification7: C07K 19/00, 14/005, C12N 5/10, 15/33
- (21) International Application Number: PCT/EP01/12975
- (22) International Filing Date:
 9 November 2001 (09.11.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 00124629.7 10 November 2000 (10.11.2000) E 01109543.7 17 April 2001 (17.04.2001) E 60/311,876 13 August 2001 (13.08.2001) U
- (71) Applicant: ARTEMIS PHARMACEUTICALS GMBH [DE/DE]; Neurather Ring 1, 51063 Cologne (DE).
- (72) Inventors: KÜHN, Ratf; Hellenthaler Str. 22, 50937 Köln (DE). FELDER, Susanne; Kopenhagener Str. 1, 41469 Neuss (DE). SCHWENK, Frieder; Konradenweg 15, 50999 Köln (DE). KÜTER-LUKS, Birgit; Niederichstr. 10, 50668 Köln (DE). FAUST, Nicole; Dammelsfurther Weg 9, 51063 Rösrath (DE).
- (74) Agents: HELBING, Jörg et al.; Von Kreisler Selting Werner, Postfach 10 22 41, 50462 Köln (DE).

- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report
- entirely in electronic form (except for this front page) and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

A2

(54) Title: MODIFIED RECOMBINASE

(57) Abstract: The present invention concerns a fusion protein comprising a recombinase protein, preferably the site-specific DNA recombinase C31-Int of phage (C31, and a peptide sequence which directs the nuclear uptake of the fusion protein in eucaryotic cells, and the use of this fusion protein to recombine, invert or delete DNA molecules containing recognition sequences for said recombinase in eucaryotic cells at high efficiency. In addition the invention relates to a cell, preferably a mammalian cell which contains recognition sequences for said recombinase in its genome and wherein the genome is recombined by the action of said fusion protein. Moreover, the invention relates to the use of said cell to study the function of genes and for the creation of transgenic organisms to study gene function at various developmental stages, including the adult. In conclusion, the present invention provides a process which enables the highly efficient modification of the genome of mammalian cells by site-specific recombination.

Modified Recombinase

The present invention concerns a fusion protein comprising a recombinase 5 protein, preferably the site-specific DNA recombinase C31-Int of phage Φ C31, and a peptide sequence which directs the nuclear uptake of the fusion protein in eucaryotic cells, and the use of this fusion protein to recombine, invert or delete DNA molecules containing recognition sequences for said recombinase in 10 eucaryotic cells at high efficiency. In addition the invention relates to a cell, preferably a mammalian cell which contains recognition sequences for said recombinase in its genome and wherein the genome is recombined by the action of said fusion protein. Moreover, the invention relates to the use of said cell to study the function of genes and for the creation of transgenic organisms to study 15 gene function at various developmental stages, including the adult. In conclusion, the present invention provides a process which enables the highly efficient modification of the genome of mammalian cells by site-specific recombination.

20 Background of the invention

The controlled and permanent modification of the genome of eucaryotic cells and organisms is an important method for research applications, e.g. for studying gene function, for medical applications like gene therapy and the creation of disease models and for the design of economically important animals and crops. The basic methods for genome manipulations by the engineering of endogenous 25 genes through gene targeting in murine embryonic stem (ES) cells are well established and used since many years (Capecchi, Trends in Genetics, 5, 70-76 (1989)). Since ES cells can pass mutations induced in vitro to transgenic offspring in vivo it is possible to analyse the consequences of gene disruption in the context of the entire organism. Thus, numerous mouse strains with functionally inactivated genes ("knock-out mice") have been created by this technology and utilised to study the biological function of a variety of genes (Koller et al., Ann. Rev. Immunol., 10, 705 - 730 (1992)). More importantly, mouse mutants created by this procedure (also known as "conventional, complete or classical mutants"), contain the inactivated gene in all cells and 35

1

10

15

20

25

30

tissues throughout life. Thus, classical mouse mutants represent the best animal model for inherited human diseases as the mutation is introduced into the germline but are not the optimal model to study gene function in adults, e.g. to validate potential drug target genes.

A refined method of targeted mutagenesis, referred to as conditional mutagenesis, employs the Cre/loxP site-specific recombination system which enables the temporally and/or spatially restricted inactivation of target genes in cells or mice (Rajewsky et al., J. Clin. Invest., 98, 600 - 603 (1996)). The phage P1 derived Cre recombinase recognises a 34 bp sequence referred to as loxP site which is structured as an inverted repeat of 13 bp separated by an asymmetric 8 bp sequence which defines the direction of the loxP site. If two loxP sites are located on a DNA molecule in the same orientation the intervening DNA sequence is excised by Cre recombinase from the parental molecule as a closed circle leaving one loxP site on each of the reaction products (Kilby et al., TIG, 9, 413-421 (1993)). The creation of conditional mouse mutants initially requires the generation of two mouse strains, one containing two or more Cre recombinase recognition (loxP) sites in its genome while the other harbours a Cre transgene. The former strain is generated by homologous recombination in ES cells as described above, except that the exon(s) of the target gene is (are) flanked by two loxP sites which reside in introns and do not interfere with gene expression. The Cre transgenic strain contains a transgene whose expression is either constitutively active in certain cells and tissues or is inducible by external agents, depending on the promoter region used. Crossing of the loxP-flanked mouse strain with the Cre recombinase expressing strain enables the deletion of the loxP-flanked exons in the genome of doubly transgenic offspring in a prespecified temporally and/or spatially restricted manner. Thus, the method allows the analysis of gene function in particular cell types and tissues of otherwise widely expressed genes. Moreover, it enables the analysis of gene function in the adult organism by circumventing embryonic lethality which is often the consequence of complete (germline) gene inactivation. For pharmaceutical research, aiming to validate the utility of genes and their products for drug development, gene . inactivation which is inducible in adults provides an excellent genetic tool as this mimicks the biological effects of target inhibition upon drug application.

Since the first description of the concept of conditional gene targeting using the Cre/loxP system in mice in 1994 (Gu et al., Science 265, 103-106 (1994)) this method became increasingly popular among the research community and resulted in a broad collection of genetic tools for biological research in the mouse. More than 30 Cre transgenic mouse strains with various tissue specificities for gene inactivation have been created, including several "deleter" strains which allow to remove the loxP-flanked target gene segment in the male or female germline (Cohen-Tannoudji et al., Mol. Hum. Reprod. 4, 929-938 (1998); Metzger et al., Curr. Op. Biotech., 10, 470-476 (1999)). The need to characterise the expression pattern of Cre mediated recombination in newly generated strains stimulated the construction of a number of "Cre-reporter" strains which harbour a silent reporter gene the expression of which is activated upon Cre-mediated deletion (Nagy, Genesis, 26, 99-109 (2000)). Conditional mouse mutants have been reported for about 20 different genes, many of them could not be studied in adults as their complete inactivation leads to embryonic lethality (Cohen-Tannoudji et al., Mol. Hum. Reprod. 4, 929-938 (1998)).

10

15

20

25

30

35

Great efforts have also been made to control the expression of Cre recombinase in an inducible fashion in mice. After the first demonstration that inducible gene knock-out is feasible in adult mice using an interferon controlled promoter (Kühn et al., Science, 269, 1427-1429 (1995)), mainly two methods were applied to control the activity of Cre recombinase. First, it has been demonstrated that the fusion of Cre with the ligand binding domain of a mutant estrogen receptor allows to control recombinase activity by a specific steroid-like inducer. Several transgenic mouse strains expressing such a fusion protein have been generated and allow to induce gene inactivation in specific tissues (Metzger et al., Curr. Op. Biotech., 10, 470-476 (1999)). Furthermore, the tetracycline-regulated gene expression system has been successfully used to control the expression of Cre in transgenic mice and thus provides a second system for inducible gene inactivation using doxycycline as inducer (Saam et al., J. Biol. Chem. 274, 38071-38082 (1999)).

In addition to the application of Cre/loxP for gene inactivation by deletion of a gene segment this recombination system has been proved to be useful also for a number of other genomic manipulations in ES cells or mice. These include the

conditional activation of transgenes in mice, chromosome engineering to obtain deletion, translocation or inversion, the simple removal of selection marker genes, gene replacement, the targeted insertion of transgenes and the (in)activation of genes by inversion (Nagy, Genesis, 26, 99-109 (2000); Cohen-Tannoudji et al., Mol. Hum. Reprod. 4, 929-938 (1998)). In conclusion, the Cre/loxP recombination system has been proven to be extremely useful for the analysis of gene function in mice by broadening the methodological spectrum for genome engineering. It can be expected that many of the protocols now established for the mouse may be applied in future also to other animals or plants.

5

10

15

20

25

30

35

In contrast to the huge diversity of genome manipulations which have been developed for the Cre/loxP system, very limited efforts have been made to develop further site-specific recombination systems for the use in mammalian cells. Alternative recombination systems of different specificity but with an efficiency comparable to Cre/loxP could further enhance the flexibility of genome engineering by the side to side use of two or more systems in the same cell or organism. Furthermore, unidirectional recombination systems which follow a different mechanism than the reversible Cre/loxP-mediated recombination should allow to develop new applications for genome engineering which cannot be performed with the current systems.

The reasons for the almost exclusive use of the Cre/loxP system for site-specific recombination in mammalian cells are readily explained by a number of requirements which must be fulfilled for the efficient use of a recombinase in mammalian cells:

- i) the recombinase should act independent of cofactors like helper proteins,
- ii) it should act independent of the supercoiling status of the target DNA and also on mammalian chromatin,
 - iii) it should be efficiently active and stable at a temperature of 37°C, and
 - it should recognize a target sequence which is sufficiently long to be unique among large genomes, and it should exhibit a very high affinity to its target site for efficient action (Kilby et al., TIG, 9, 413-421 (1993)).

Among the more than 200 described members of the integrase and resolvase/invertase recombinase families only the Cre/loxP system is presently known to fulfill all of these requirements (Nunes-Düby et al., Nucleic Acids Res., 26, 391-406 (1998); Kilby et al., TIG, 9, 413-421 (1993); Ringrose et al., J. Mol. 5 Biol., 284, 363 - 384 (1998)). Besides Cre/loxP a few recombinases have been shown to exhibit some activity in mammalian cells but their practical value is presently unclear as their efficieny has not been compared to the Cre/loxP system on the same genomic recombination substrate and in some cases it is known that one or more of the criteria listed above are not met. The best 10 characterised examples are the yeast derived FLP and Kw recombinases which exhibit a temperature optimum at 30°C but which are unstable at 37°C (Buchholz et al., Nature Biotech., 16, 657 - 662 (1998); Ringrose et al., Eur. J. Biochem., 248, 903 - 912). For FLP it has been shown in addition that its affinity to the FRT target site is much lower as compared to the affinity of Cre to loxP sites 15 (Ringrose et al., J. Mol. Biol., 284, 363 - 384 (1998)). Other recombinases which show in principle some activity in mammalian cells are a mutant integrase of phage $\lambda,$ the integrases of phages $\Phi C31$ and HK022, mutant $\gamma\delta\text{-resolvase}$ and $\beta\text{-}$ recombinase (Lorbach et al., J. Mol. Biol., 296, 1175 - 81 (2000); Groth et al., Proc. Natl. Acad. Sci. USA, 97, 5995 - 6000 (2000); Kolot et al., Mol. Biol. Rep. 20 26, 207 - 213 (1999); Schwikardi et al., FEBS Lett., 471, 147 - 150 (2000); Diaz et al., J. Biol. Chem., 274, 6634 - 6640 (1999)). Other phage integrase systems include coliphage P4 recombinase, Listeria phage recombinase, bacteriophage R4 Sre recombinase, CisA recombinase, XisF recombinase and transposon Tn4451 TnpX recombinase (Stark et al. Trends in Genetics 8, 432-439 (1992); Hatfull & 25 Gridley, in Genetic Recombination. Eds. Kucherlipati & Smith, Am. Soc. Microbiol., Washington DC, 357-396 (1988)).

However, the practical value of these recombinases and integrases for use in mammalian cells is limited as their efficiency to recombine mammalian genomic DNA has not been tested or compared with the Cre/loxP system. From the data available it can be assumed that these recombinases are much less effective than the Cre/loxP system.

In a few cases attempts have been made to improve the performance of recombinases in mammalian cells: for FLP a mutant showing improved thermostability and acticity at 37°C has been isolated but this mutant is still considerably more heat labile as compared to Cre (Buchholz et al., Nature Biotech., 16, 657 – 662 (1998)). In the case of λ -integrase and $\gamma\delta$ -resolvase the absolute requirement for coproteins and supercoiled DNA could be eliminated by the introduction of specific point mutations (Schwikardi et al. FEBS Lett 471, pp147-50 (2000)).

The import of cytoplasmic proteins into the nucleus of eucaryotic cells through nuclear pores is a regulated, energy dependent process mediated by specific receptors (Görlich et al., Science, 271, 1513 – 1518 (1996)). Proteins which do not posses a signal sequence recognised by the nuclear import machinery are excluded from the nucleus and remain in the cytoplasm. Numerous of such nuclear localisation signal sequences (NLS), which share a high proportion of basic amino acids in common, have been characterised (Boulikas, Crit. Rev. Eucar. Gene Expression, 3, 193 – 227 (1993)), the prototype of which is the 7 amino acid NLS derived from the T-antigen of the SV40 virus (Kalderon et. al, Cell, 39, 499 – 509 (1984)).

20

25

30

35

5

10

15

It was believed that the fusion of such an NLS peptide to a recombinase possibly would enhance the efficiency of the recombinase by mediating its import into the nucleus and therewith increasing the concentration of the recombinase inside the nucleus. However, for Cre recombinase it has been shown that the addition of the SV-40 T-antigen NLS does not improve its recombination efficiency in mammalian cells (Le et al., Nucleic Acid Res., 27, 4703 –4709 (1999)). Nevertheless, both Cre and a Cre-NLS-fusion protein are widely used. Schwikardi (Schwikardi et al., FEBS Lett. 471, pp147-50 (2000)) reported a $\gamma\delta$ -resolvase-SV-40 T-antigen NLS fusion protein, which also did not enhance the recombination efficiency.

The level of activity exhibited by recombinases of diverse prokaryotic origin in mammalian cells may be the result of the intrinsic properties of an enzyme depending on parameters like its temperature optimum, its target site affinity, protein structure and stability, the degree of cooperativity, the stability of the

synaptic complex and the dependence on coproteins or supercoiled DNA. Within the specific environment of mammalian cells the activity of a prokaryotic recombinase could be limited by additional factors such as a short half-life of the recombinase transcript, a short half-life of its protein, its inability to act on histone-complexed and higher order structured mammalian genomic DNA, exclusion from the nucleus or the recognition of cryptic splice sites within its mRNA resulting in a nonfunctional transcript. Due to the lack of information on the parameters listed above for almost all recombinases it is presently not possible to rationally optimise their performance in mammalian cells.

10

15

20

5

Summary of the Invention

The object to be solved by the invention of the present application is the provision of a recombination system alternative to the Cre/loxP system, which has a different specificity but an efficiency comparable to Cre/loxP. Such an alternative recombination system is particularly desirable for all those applications which require more than one potent recombination system for being successfully carried out (e.g. the methods disclosed in PCT/EP01/00060 and PCT/EP00/10162). Most surprisingly, it was found that the above object can be solved by fusing a signal peptide capable directing the nuclear import (hereinafter shortly referred to as nuclear localisation signal sequences (NLS)) to specific recombinases.

In contrast to the wildtype recombinases, the resulting modified recombinases allow a highly efficient recombination of extrachromosomal and chromosomal DNA in mammalian cells, and a highly efficient excision of extrachromosomal and chromosomal DNA-stretches, which are flanked by suitable recognition sites for said modified recombinases.

- 30 The present invention thus provides:
 - (1) A fusion protein (hereinafter also referred to as "modified recombinase") comprising
 - (a) a recombinase domain comprising a recombinase protein or fragment thereof and
- 35 (b) a signal peptide domain being linked to (a) and directing the nuclear import

of said fusion protein in eucaryotic cells,

preferably the activity of the fusion protein in eucaryotic cells is significantly higher as compared to the acitivity of the wildtype recombinase corresponding to the recombinase of the recombinase domain;

- 5 (2) in a preferred embodiment of the fusion protein defined in (1) above, the recombinase domain comprises an integrase protein, preferably a phage ΦC31 integrase (C31-Int) protein or a mutant thereof;
 - (3) a DNA coding for the fusion protein as defined in (1) or (2) above;
 - (4) a vector containing the DNA as defined in (3) above;
- 10 (5) a microorganism containing the DNA of (3) above and/or the vector of (4) above;
 - (6) a process for preparing the fusion protein as defined in (1) or (2) above which comprises culturing a microorganism as defined in (5) above;
- (7) the use of the fusion protein as defined in (1) or (2) above to recombine DNA
 molecules, which contain recombinase recognition sequences for the recombinase protein of the recombinase domain, in eucaryotic cells;
 - (8) a cell, preferably a mammalian cell containing the DNA sequence of (3) above in its genome;
- (9) the use of the cell of (8) above for studying the function of genes and for thecreation of transgenic organisms;
 - (10) a transgenic organism, preferably a transgenic mammal containing the DNA sequence of (3) above in its genome;
 - (11) the use of the transgenic organism of (10) above for studying gene function at various developmental stages; and
- 25 (12) a method for recombining DNA molecules of cells or organisms containing recognition sequences for the recombinase protein of the recombinase domain as defined in (1) or (2) above, which method comprises supplying the cells or organisms with a fusion protein as defined in (1) or (2) above, or with a DNA sequence of (3) above and/or a vector of (4) above which are capable of expressing said fusion protein in the cell or organism.

The present invention combines the use of prokaryotic recombinases such as the C31-Int with a eukaryotic signal sequence which increases its efficiency in mammalian cells such that it is equal to the widely used Cre/loxP recombination system. The improved recombination system of the present invention provides

35

an alternative recombination system for use in mammalian cells and organisms which allows to perform the same types of genomic modifications as shown for Cre/loxP, including conditional gene inactivation by recombinase-mediated deletion, the conditional activation of transgenes in mice, chromosome engineering to obtain deletion, translocation or inversion, the simple removal of selection marker genes, gene replacement, the targeted insertion of transgenes and the (in)activation of genes by inversion.

Short Description of Figures

vectors into CHO cells.

25

- 10 <u>Fig. 1:</u> C31-Int and Cre recombinase expression vectors and a recombinase reporter vector used for transient and stable transfections
 - <u>Fig. 2:</u> Results of transient transfections of C31 Int and Cre expression vectors and reporter vectors into CHO cells.
- Fig. 3: Results of transient transfections of XisA and Ssv recombinase expression vectors with and without nuclear localisation signals and reporter
- Fig. 4: Results of transient transfections of C31-Int and Cre recombinase vectors into a stable reporter cell line.
 - Fig. 5: In situ detection of β -galactosidase in 3T3(pRK64)-3 cells transfected with recombinase expression vectors
 - Fig. 6: Test vector for C31-Int mediated deletion, pRK64, and the expected deletion product.
- Fig. 7: PCR products generated with the primers P64-1 and P64-4 and sequence comparison.
 - Fig. 8: ROSA26 locus of the C31 reporter mice carrying a C31 reporter construct.
- Fig. 9: In situ detection of β -galactosidase in a cryosection of the testis of: (A) a double transgenic mouse carrying both the recombinase and the reporter; and

WO 02/38613 PCT/EP01/12975 10

(B) a transgenic mouse carrying only the reporter as a control.

Detailed Description of the Invention

20

25

30

35

The "organisms" according to the present invention are multi-cell organisms and can be vertebrates such as mammals (humans and non-human animals including rodents such as mice or rats) or non-mammals (e.g. fish), or can be invertebrates such as insects or worms, or can be plants (higher plants, algi or fungi). Most preferred living organisms are mice and fish.

10 "Cells" and "eucaryotic cells" according to the present invention include cells isolated from the above defined living organism and cultured in vitro. These cells can be transformed (immortalized) or untransformed (directly derived from the living organism; primary cell culture).

15 "Microorganism" according to the present invention relates to procaryotes (e.g. E. coli) and eucaryotic microorganisms (e.g. yeasts).

According to embodiment (1) of the present invention, the activity of the fusion protein in eucaryotic cells is significantly higher as compared to the acitivity of the wildtype recombinase corresponding to the recombinase of the recombinase domain. A "significantly higher activity" in accordance with the present invention refers to an increase in activity of at least 50%, preferably at least 75%, more preferably at least 100% relative to the corresponding wildtyp recombinase in eucaryotic cells. A "significantly higher activty" also implies that the resulting fusion protein has at least 25%, preferably at least 50% and more preferably at least 75%, of the activity of Cre/loxP in 3T3 cells with a stably integrated target sequence.

Recombinase proteins which can be used in the recombinase domain of the fusion protein of the present invention (i.e., giving a fusion having a "significantly higher activty" as defined above) include, but are not limited to, a certain type of recombinases belonging to the family of of large serine recombinases (Thorpe et al., Control of directionalty in the site-specific recombination system of the streptomyces phage φC31, Molecular Microbiology 38(2), 232-241 (2000)). This family includes bacteriophage ΦC31 integrase ("C31-Int"; the amino acid WO 02/38613 PCT/EP01/12975

sequence of said integrase and a DNA sequence coding therefor are shown in SEQ ID NOs:21 and 20, respectively), coliphage P4 recombinase, Listeria phage recombinase, bacteriophage R4 Sre recombinase ("R4 Sre" deposited under GI 793758; the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:55 and 54, respectively), bacillus subtilis CisA recombinase ("CisA" deposited under GI 142689; the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:57 and 56, respectively), XisF recombinase from annabaena sp. Strain PCC 7120 (Cyanobacterium; "XisF" deposited under GI 349678; the amino acid sequence of said integrase and a DNA sequence coding therefor are shown in SEQ ID NOs:59 and 58, respectively), transposon Tn4451 TnpX recombinase ("TnpX" deposited under GI 551135; the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:61 and 60, respectively), "XisA" recombinase from annabaena sp. Strain PCC 7120 (Cyanobacterium; the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:63 and 62, respectively), "SSV" recombinase from phage of sulfolobus shibatae (the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:65 and 64, respectively), lactococcal bacteriophage TP901-1 recombinase (TP901-1 complete genome deposited under GI 13786531; the amino acid sequence of said recombinase and a DNA sequence coding therefor are shown in SEQ ID NOs:108 and 107, respectively), and the like, or mutants thereof. Other procaryotic recombinases known in the art are also applicable.

10

15

20

A "mutant" of the above recombinases in accordance with the present invention relates to a mutant of the respective original (viz. wild-type) recombinase having a recombinase activity similar (e.g. at least about 90%) to that of said wild-type recombinase. Mutants include truncated forms of the recombinase (such as N- or C-terminal truncated recombinase proteins), deletion-type mutants (where one or more amino acid residues or segments having more than one continuous amino acid residue have been deleted from the primary sequence of the wildtyp recombinase), replacement-type mutants (where one or more amino acid residues or segments of the primary sequence of the wildtyp recombinase have been replaced with alternative amino acid residues or segments), or combinations thereof.

According to embodiment (2) of the invention, the recombinase domain comprises an integrase protein, preferably a phage Φ C31 integrase (C31-Int) protein or a mutant thereof. Thus, the present invention provides a fusion protein comprising

- (a) an integrase domain being a C31-Int protein or a mutant thereof, and (b) a signal peptide domain being linked to (a) and directing the nuclear import of said fusion protein into eucaryotic cells.
- In the fusion protein of embodiment (2), the integrase domain is preferably a C31-Int having the amino acid sequence shown in SEQ ID NO:21 or a C-terminal truncated form thereof. Suitable truncated forms of the C31-Int comprise amino acid residues 306 to 613 of SEQ ID NO:21.
- The signal peptide domain (hereinafter also referred to as "NLS") is preferably derived from yeast GAL4, SKI3, L29 or histone H2B proteins, polyoma virus large T protein, VP1 or VP2 capsid protein, SV40 VP1 or VP2 capsid protein, Adenovirus E1a or DBP protein, influenza virus NS1 protein, hepatitis virus core antigen or the mammalian lamin, c-myc, max, c-myb, p53, c-erbA, jun, Tax, steroid receptor or Mx proteins (see Boulikas, Crit. Rev. Eucar. Gene Expression, 3, 193 227 (1993)), simian virus 40 ("SV40") T-antigen (Kalderon et. al, Cell, 39, 499 509 (1984)) or other proteins with known nuclear localisation. The NLS is preferably derived from the SV40 T-antigen.
- Furthermore, the signal peptide domain preferably has a length of 5 to 74, preferably 7 to 15 amino acid residues. More preferred is that the signal peptide domain comprises a segment of 6 amino acid residues wherein at least 2 amino acid residues, preferably at least 3 amino acid residues are positively charged basic amino acids. Basic amino acids include, but are not limited to, Lysin, Arginin and Histidine. Particularly preferred signal peptides are show in the following table.

| | Organism | Sequence/(SEQ ID N | 10:) |
|----|------------|------------------------------------|------|
| | yeast GAL4 | MKx11CRLKKLKCSKEKPKCAKCLKx5Rx3KTKR | (24) |
| 35 | yeast SKI3 | IKYFKKFPKD | (25) |

| | WO 02/38613 | | PCT/EP01/12975 | |
|----|-----------------------------------|----|-----------------------|-------|
| | yeast L29 | 13 | MTCCVTDVUDCCCA | (0.0) |
| | , 3.33 223 | | MTGSKTRKHRGSGA | (26) |
| | yeast histone H2B | | (MTGSKHRKHPGSGA) | (27) |
| | | | (GKKRSKA) | (28) |
| _ | polyoma virus large T protein | | (PKKAREDVSRKRPR) | (29) |
| 5 | polyoma virus VP1 capsid protein | | (APKRKSGVSKC) | (30) |
| | polyoma virus VP2 capsid protein | | (EEDGPQKKKRRL) | (31) |
| | SV40 VP1 capsid protein | | (APTKRKGS) | (32) |
| | SV40 VP2 capsid protein, | | (PNKKKRK) | (33) |
| | Adenovirus E1a protein | | (KRPRP) | (34) |
| 10 | | | (CGGLSSKRPRP) | (35) |
| | Adenovirus DBP protein | | (PPKKRMRRRIEPKKKKKRP) | (36) |
| | influenza virus NS1 protein | | (PFLDRLRRDQK) | (37) |
| | · | | (PKQKRKMAR) | (38) |
| • | human laminA | | (SVTKKRKLE) | (39) |
| 15 | human c-myc | | (CGGAAKRVKLD) | (40) |
| | | | (PAAKRVKLD) | (41) |
| | | | (RQRRNELKRSP) | (42) |
| | HUMAN max | | (PQSRKKLR) | (43) |
| | HUMAN c-myb | | (PLLKKIKQ) | (44) |
| 20 | HUMAN p53 | | (PQPKKKP) | (45) |
| | HUMAN c-erbA | | (SKRVAKRKL) | (46) |
| | VIRAL jun | | (ASKSRKRKL) | (47) |
| | HUMAN Tax | | (GGLCSARLHRHALLAT) | (48) |
| | Mammalian glucocorticoid receptor | | (RKTKKKIK) | (49) |
| 25 | HUMAN ANDROGEN RECEPTOR | | (RKLKKLGN) | (50) |
| | MAMMALIAN ESTROGEN RECEPTOR | | (RKDRRGGR) | (51) |
| | Mx proteins | | (DTREKKKFLKRRLLRLDE) | (52) |
| | SV40 T-antigen | | (PKKKRKV) | (53) |
| | | | • | ` - / |

35

The most preferred signal peptide domain is that of SV40 T-antigen having the 30 sequence Pro-Lys-Lys-Lys-Arg-Lys-Val.

The signal peptide domain may be linked to the N-terminal or C-terminal of the integrase domain or may be integrated into the integrase domain, preferably the signal peptide domain is linked to the C-terminal of the integrase domain. With

regard to phage Φ C31 integrase protein of embodiment (2) of the invention it was found that the fusion of an NLS-peptide to the C-terminus of the integrase provided a much higher increase of activity as compared to the fusion of the same NLS-peptide to the N-terminus of the integrase (see Example 1, figures 3 and 4).

According to the present invention, the signal peptide domain may be linked to the integrase domain directly or through a linker peptide. Suitable linkers include peptides having from 1 to 30, preferably 1 to 15 amino acid residues, said amino acid residues being essentially neutral amino acids such as Gly, Ala and Val.

The most preferred fusion protein of the present invention comprises the amino acid sequence shown in SEQ ID NO:23 (a suitable DNA sequence coding for said fusion protein being shown in SEQ ID NO:22).

15

20

25

30

35

10

Further preferred fusion proteins of the present invention are "NLS-XisA" and "NLS-SSV" (having the NLS-peptide fused to the N-terminus of the recombinases) as shown in SEQ ID NO:67 and 69, respectively (suitable DNA sequences coding for said fusion proteins being shown in SEQ ID NO:66 and 68, respectively).

In embodiments (7), (8), (10) and (12) of the invention the DNA molecules, the cell or transgenic organism may also contain recognition sequences for the recombinase protein of the recombinase domain. Thus, when utilizing the fusion protein of embodiment (2), the C31-Int recognition sequences attP and attB are present in DNA molecules, the cell or transgenic organism.

The term "mammal" as used in embodiment (10) of the invention includes non-human mammals (viz. animals as defined above) and humans (if such subject matter is patentable with the respective patent authority).

Since the modified recombinase of the invention, in particular the modified C31-Int, acts in mammalian cells as efficient (or at least almost as efficient) as the widely used Cre/loxP system it can be used for a large variety of genomic modifications (including the methods disclosed in PCT/EP01/00060 and

PCT/EP00/10162, the content of which is herewith incorporated by reference). Concerning embodiment (11) it is to be noted that the mammals of embodiment (10) can be used to study the function of genes, e.g. in mice, by conditional gene targeting. For this purpose suitable recognition sequences - when utilizing the fusion protein of embodiment (2), one attP and one attB site (C31-Int recognition sequences) in the same orientation - can be introduced into introns of a gene by homologous recombination of a gene targeting vector in ES cells such that the two sites flank one or more exons of the gene to be studied but do not interfere with gene expression. A selection marker gene, needed to isolate recombinant ES cell clones, can be flanked by two recognition sites of another recombinase such as loxP or FRT sites to enable deletion of the marker gene upon transient expression of the respective recombinase in ES cells. These ES cells can be used to generate germline chimaeric mice which transmit the target gene modified by att sites to their offspring and allow to establish a modified mouse strain. The crossing of this strain with a C31-Int recombinase transgenic line or the application of C31-Int protein will result in the deletion of the att-flanked gene segment from the genome of doubly transgenic offspring and the inactivation of the target gene in doubly transgenic offspring in a prespecified temporally and/or spatially restricted manner. The C31-Int transgenic strain contains a transgene whose expression is either constitutively active in certain cells and tissues or is inducible by external agents, depending on the promoter region used. If an attB and an attP site are placed into the genome in opposite orientation C31-Int mediated recombination results in the irreversible inversion of the flanked gene segment leading the functional loss of on or more exons of the target gene. Thus, the method allows the analysis of gene function in particular cell types and tissues of otherwise widely expressed genes and circumvents embryonic lethality which is often the consequence of complete (germline) gene inactivation. For the validation of genes and their products for drug development, gene inactivation which is inducible in adults provides an excellent genetic tool as this mimicks the biological effects of target inhibition upon drug application. If a pair of attB/P sites is placed in the same or opposite orientation into a chromosome at large distance using two gene targeting vectors, C31-Int recombination allows to delete or invert chromosome segments containing one or more genes, or chromosomal translocations if the two sites are located on different chromosomes. In another application of the method a pair of attB/P sites is

10

15

20

25

30

35

PCT/EP01/12975

placed in the same orientation within a transgene such that the deletion of the att-flanked DNA segment results in gene expression, e.g. of a toxin or reporter gene for cell lineage studies, or in the inactivation of the transgene.

In addition, according with embodiment (12) of the invention, the recombination system of embodiment (1), in particular the C31-Int recombination system of embodiment (2), can also be used for the site specific integration of foreign DNA into the genome of mammalian cells, e.g. for gene therapy. For this purpose, and if the C31-Int recombination system of embodiment (2) is utilized, only one attB (or attP) site is initially introduced into the genome by homologous 10 recombination, or an endogenous genomic sequence which resembles attB or attP is used . The application of a vector containing an attP (or attB) site to such cells or mice in conjunction with the expression of C31-Int recombinase will lead to the site specific integration of the vector into the genomic att site.

15

Thus, the present invention provides a process which enables the highly efficient modification of the genome of mammalian cells by site-specific recombination. Said process possesses the following advantages over current technology:

- the modified recombinase, in particular the modified C31-Integrase, allows 20 (i) to recombine extrachromosomal and genomic DNA in mammalian cells at much higher efficiency as compared to the use of its wildtype form;
- the modified recombinase, in particular the modified C31-Integrase, is the (ii) first described alternative recombination system with equal efficiency to 25 Cre/loxP for the recombination of chromosomal DNA in mammalian cells.

The appended figures further explain the present invention:

30 Figure 1 shows C31-Int and Cre recombinase expression vectors and a recombinase reporter vector used for transient and stable transfections. A-D: Mammalian expression vectors for recombinases which contain the CMV immediate early promoter followed by a hybrid intron, the coding region of the recombinase to be tested, and an artificial polyadenylation signal sequence (pA).

A: pCMV-C31Int(wt) containing the nonmodified (wildtype) 1.85 kb coding region of C31-Int as found in the genome of phage Φ X31.

B: pCMV-C31Int(NNLS) containing a modified C31-Int gene coding for the full length C31-Int protein with a N-terminal fusion to the SV40 virus large T antigen nuclear localisation signal (NLS).

C: pCMV-C31Int(CNLS) containing a modified C31-Int gene coding for the full length C31-Int protein with a C-terminal fusion to the SV40 virus large T antigen nuclear localisation signal (NLS).

D: pCMV-Cre contains the 1.1 kb Cre coding region with an N-terminal fusion to the SV40 T antigen NLS.

E: Recombination substrate vector pRK64 contains a SV40 promoter region followed by a 1.1 kb cassette consisting of the coding region of the puromycin resistance gene and a polyadenylation signal sequence, flanked 5` by the 84 bp attB and 3´ by the 84 bp attP recognition site of C31-Int. pRK64 contains in addition two Cre recognition (loxP) sites in direct orientation next to the att sites.

<u>Figure 2</u> shows results of transient transfections of C31-Int and Cre recombinase and reporter vectors into CHO cells.

All transfections were performed with a fixed amount of the reporter plasmid pRK64 and 0.5 ng or 1 ng of the recombinase expression plasmids pCMV-C31-Int(wt) (samples 4-5), pCMV-C31-Int(NNLS) (samples 6-7), pCMV-C31-Int(CNLS) (samples 8-9) or pCMV-Cre (samples 10-11). Negative controls: transfection with pRK64 (sample 3) or pUC19 alone (sample 1). Positive control: transfection with the Cre-recombined reporter pRK64(ΔCre) (sample 2).

The vertical rows show the mean values and standard deviation of "Relative Light Units" obtained from lysates with the assay for β-galactosidase (RLU (β-Gal)), the RLU from the assay for Luciferase, the ratio of the β-galactosidase and Luciferase values with standard deviation (RLU x 10⁵ (Gal/Luc)), and the relative activity of the various recombinases as compared to the positive control defined as 1.

<u>Figure 3</u> shows results of transient transfections of XisA and Ssv recombinases and reporter vectors into CHO cells.

All transfections were performed with fixed amounts of the reporter plasmids pPGKnif (for XisA) and pPGKattA (for SSV) and 25 ng or 100 ng of the

recombinase expression plasmids pCMV-XisA, pCMV-XisA(NNLS) and 10 ng or 20 ng of the expression plasmids pCMV-Ssv and pCMV-Ssv(NNLS). Negative controls: transfection with pPGKnif or pPGKattA alone.

The vertical rows show the mean values and standard deviation of "Relative Light Units" obtained from lysates with the assay for β -galactosidase (RLU (β -Gal)), the RLU from the assay for Luciferase, the ratio of the β -galactosidase and "Luciferase" values with standard deviation (RLU x 10^5 (Gal/Luc)).

<u>Figure 4</u> shows results of transient transfections of recombinase vectors into a stable reporter cell line.

All transfections were performed with a NIH 3T3 derived clone containing stably integrated copies of the pRK64 recombination substrate vector. Either 32 ng or 64 ng of the recombinase expression plasmids pCMV-C31-Int(wt) (samples 2-3), pCMV-C31-Int(NNLS) (samples 4-5), pCMV-C31-Int(CNLS) (samples 6-7) or pCMV-Cre(NNLS) (samples 8-9). Negative control: transfection with pUC19 alone (sample 1).

The vertical rows show the mean values and standard deviation of "Relative Light Units" obtained from lysates with the assay for β -galactosidase (RLU (β -Gal)) and the relative activity of the various recombinases as compared to the value obtained with pCMV-Cre(NNLS) defined as 1.

Figure 5 shows the in situ detection of β -galactosidase in 3T3(pRK64)-3 cells transfected with recombinase expression vectors.

The Cre and C31-Int recombinase reporter cell line 3T3(pRK64)-3 was either not transfected with DNA (A), transfected with the Cre expression vector pCMV-Cre (B) or with the C31-Int expression vector pCMV-C31-Int(CNLS). Two days after transection the cells were fixed and incubated with the histochemical X-Gal assay which develops a blue stain in β-galactosidase positive cells indicating recombinase mediated activation of the reporter gene.

30

35

10

15

20

<u>Figure 6</u> shows the test vector for C31-Int mediated deletion, pRK64, and the expected product of deletion, pRK64(Δ Int).

Plasmid pRK64 contains the 1.1 kb cassette of the coding region of the puromycin resistance gene and a polyadenylation signal, which is flanked 5' by the 84 bp attB and 3' by the 84 bp attP recognition site (large triangles) of C31-

Int. These attB and attP sites are oriented in the same way to each other (thick black arrows) which is used by the Φ X31 phage to integrate into the bacterial genome. In addition, the cassette is flanked by two Cre recombinase recognition (loxP) sites in the same orientation (black small triangles). For better orientation the half sites of the att sequences are labelled by a direction (thin arrow) and numbered 1-4. The 3 bp sequence within the att sites at which recombination occurs is framed by a box. The positions at which the PCR primers P64-1 and P64-4 hybridise to the pRK64 vector are indicated by arrows, pointing into the 3 direction of both oligonucleotides.

PRK64(ΔInt) depicts the deletion product expected from the C31-Int mediated recombination between the att sites of pRK64. The recombination between a pair of attB/attP sites generates an attR site remaining on the parental DNA molecule while the puromycin cassette is excised. In this configuration the primers P64-1 and P64-4 will amplify a PCR product of 630 bp from pRK64(ΔInt).

15

20

25

30

35

<u>Figure 7</u> shows PCR products generated with the primers P64-1 and P64-4 and a sequence comparison of the PCR product.

A: Analysis of PCR products on an agarose gel from PCR reactions using the Primers P64-1 and P64-4 on DNA extracted from MEF5-5 cells transfected 2 days before with plasmid pRK64 alone (lane 4), with pRK64 + CMV-Cre (lane 3), with pRK64 + pCMV-C31-Int(wt) (lane 2), and from a control reaction which did not contain cellular DNA (lane 1). The product with an apparent size around 650 bp, as compared to the size marker used, from lane 2 was excised from the agarose gel and purified. The PCR product was cloned into a sequencing plasmid vector and gave rise to the plasmid pRK80d. The insert of this plasmid was sequenced using reverse primer (seq80d) and compared to the predicted sequence of the pRK64 vector after C31-Int mediated deletion of the att flanked cassette, pRK64(Δ Int). The cloned PCR product shows a 100% identity with the predicted attR sequence after deletion. The generated attR site is shown in a box, with the same sequence designation used in Figure 5. The nucleotide positions (pos.) of the compared sequences pRK64(Δ Int) and Seq80d are indicated.

Figure 8 shows the modified ROSA26 locus of C31 reporter mice (Seq ID NO:106). A recombination substrate has been inserted in the ROSA26 locus. The substate consists of a splice acceptor (SA) followed by a cassette consisting of

the hygromycin resistance gene driven by a PGK promoter and flanked by the recombination sites attB and attP. In addition the reporter contains two Cre recognition sites (loxP) in direct orientation next to the att sites. This cassette is followed by the coding region for β -galactosidase, which is only expressed when the hygromycin resistance gene has been deleted by recombination.

Figure 9 shows the in situ detection of β -galactosidase activity. A cryosection of the testis of a double transgenic mouse carrying both the C31-int recombinase and the recombination substrate was stained with X-Gal (A). The blue colour indicates recombination of the substrate, which leads to the expression of β -galactosidase. As a control a cryosection of testis of a transgenic mouse carrying only the recombination substrate was stained with X-Gal (B).

The present invention is further illustrated by the following Examples which are, however, not to be construed as to limit the invention.

Examples

10

Example 1

As compared to Cre recombinase the wildtype form of C31-Int exhibits a 20 significantly lower recombination activity in mammalian cells which falls in the range of 10 - 40% of Cre, depending on the assay system used (see below). As a measure which may increase C31-Int efficiency in eukaryotic cells we designed mammalian expression vectors for N- or C-terminal fusion proteins of C31-Int with a peptide was designed which is recognised by the nuclear import 25 machinery. The recombination efficiency obtained by this modified C31-Int recombinase in mammalian cells was compared side by side to the unmodified (wildtype) form of C31-Int and to Cre recombinase. For the quantification of recombinase activities the expression vectors were transiently introduced into a mammalian cell line together with a reporter vector which contains C31-Int and 30 Cre target sites and leads to the expression of ß-galactosidase upon recombinase mediated deletion of a vector segment flanked by recombinase recognition sites.

A. Plasmid constructions:

Construction of the recombination test vectors pPGKnif and pPGKattA: first a nifD site (Haselkorn, Annu Rev.Genet. 26, 113-130 (1992)) generated by the annealing of the two synthetic oligonucleotides nifD3 (SEQ ID NO:89) and nifD4 (SEQ ID NO:90), was ligated into the BamHI restriction site of the vector PSV-Pax1 (Buchholz et al., Nucleic Acids Res., 24, 4256-4262 (1996)), 3'of its puromycin resistance gene and loxP site, giving rise to plasmid pPGKnifD3' (SEQ ID NO:79). Next, another nifD site, generated by the annealing of the two synthetic oligonucleotides nifD1 (SEQ ID NO:87) and nifD2 (SEQ ID NO:88), was ligated into the BstBI restriction site of plasmid pPGKnifD3', upstream of the puromycin resistance gene and loxP site, giving rise to plasmid pPGKnifD (SEQ ID NO:78). For pPGKattA (Muskhelishvili et al., Mol.Gen.Genet. 237, 334-342 (1993)) first a 352bp-fragment was amplified from genomic DNA from the thermophilic bacterium Sulfolobus shibatae (DSM-5389, DSMZ Braunschweig-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder 15 Weg 1b, D-38124 Braunschweig, Germany) with oligonucleotides SSV5 (SEQ ID NO:96) and SSV6 (SEQ ID NO:97) including restriction sites for BamHI and BstBI. The amplified fragment was cloned into the BamHI site of the vector PSV-Pax1 giving rise to plasmid pPGKattA1 (SEQ ID NO:82), subsequently the same 352 bp-fragment was cloned into the BstBI site of pPGKattA1 giving rise to the plasmid pPGKattA2 (SEQ ID NO:83). The sequence and orientation of both nifD sites and attA sites was confirmed by DNA sequence analysis. In pPGKnifD/pPGKattA2 the newly cloned nifD/attA sites (positions 535-619 and 1722-1787/ positions 6718-7081 and 12-363) are in the same orientation flanking the puromycln resistance gene and the SV40 early polyadenylation sequence. The nifD/attA sites are followed by loxP sites in the same orientation (positions 623 - 656 and 1794 - 1827/ positions 7085-7118 and 369-402). The puromycin cassette is transcribed from the SV40 early enhancer/promoter region and followed by the coding region for E. coli β -galactosidase and the SV40 late region polyadenylation sequence.

30

35

10

20

25

Construction of XisA and SSV expression vectors: First the XisA gene of cyanobacterium PCC7120 was amplified by PCR from genomic DNA from Nostoc strain PCC7120 (CNCM-Collection Nationale de Cultures de Microorganismes, Institut Pasteur, Paris) using the primers XisA1 (SEQ ID NO:84) and XisA3 (SEQ ID NO:86), and XisA1 (SEQ ID NO:84) and XisA2 (SEQ ID NO:85) (with NLS).

The ends of the PCR product were digested with NotI and the product was ligated into plasmid pBluescript II KS, opened with NotI, giving rise to plasmids pRK42a and pRK43 (with NNLS). The DNA sequence of the insert was determined and found to be identical to the published XisA sequence (Genbank GI:3953452) apart from four silent point mutations. The XisA gene was isolated as a 1.4 kb fragment from pRK42a and pRK43 by digestion with NotI and ligated into the generic mammalian expression vector pRK50 (see below), opened with NotI, giving rise to the XisA expression vectors pCMV-XisA (SEQ ID NO:76) and pCMV-XisA(NNLS) (SEQ ID NO:77). pCMV-XisA(wt) contains a Cytomegalovirus 10 immediated early gene promoter (position 1 - 616), a 240 bp hybrid intron (position 716 - 953), the XisA gene (position 974 - 2392), and a synthetic polyadenylation sequence (position 2413 - 2591).

5

15

20

25

30

35

The SSV gene was amplified from genomic DNA from the thermophilic bacterium Sulfolobus shibatae (DSM-5389, DSMZ Braunschweig- Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-38124 Braunschweig, Germany) in two PCR steps because of an internal attP sequence. First, two overlapping PCR fragments were created with the oligonucleotides SSV1-1 (SEQ ID NO:91) (or SSV1-2 for the SSV(NNLS) gene) and SSV2 (SEQ ID NO:93) and oligonucleotides SSV3 (SEQ ID NO:94) and SSV4 (SEQ ID NO:95).

Using these overlapping fragments as template, a 1000bp fragment containing the complete SSV coding sequence was amplified with primers SSV1-1 (or SSV1-2 for the SSV(NNLS) gene) and SSV4. The 5' 620 bp-fragments of these PCR products were isolated by digestion with NotI-XhoI and cloned into vector pBluescript II KS giving rise to plasmids pRK47 and pRK48 (with NLS). The 3' 380 bp fragment generated by XhoI-digestion was cloned into the XhoI restriction site of vector pBluescript II KS giving rise to the plasmid pBS-SSVs (SEQ ID NO:72). The 380bp SSV-fragment was then isolated by digestion of pBS-SSVs with XhoI and ligated into pRK47 and pRK48 opened by XhoI giving rise to plasmids pBS-SSV3 (SEQ ID NO:70) and pBS-SSV4 (SEQ ID NO:71) (with NLS) containing the complete SSV gene. Sequencing of the plasmids confirmed one point mutation in both plasmids. Therefore 312 bp/ 91 bp fragments generated by digestion with EcoRV-SmaI/ EcoRV-XhoI of another clone of pRK47

by sequencing. The SSV gene was isolated from pRK47 and pRK48 by digestion with NotI and KpnI and ligated into the generic mammalian expression vector

were exchanged in plasmids pBS-SSV3/ pBS-SSV4. Sequences were confirmed

pRK50 (see below), opened with NotI and SalI, giving rise to the SSV expression vectors pCMV-SSV(wt) (SEQ ID NO:74) and pCMV-SSV(NNLS) (SEQ ID NO:75).

5

10

15

20

Construction of the recombination test vector pRK64: first an attB site (Thorpe et al. Proc. Natl. Acad. Sci. USA, 95, 5505 - 5510 (1998)), generated by the annealing of the two synthetic oligonucleotides C31-4 (SEQ ID NO:1) and C31-5 (SEQ ID NO:2), was ligated into the BstBI restriction site of the vector PSV-Pax1 (Buchholz et al., Nucleic Acids Res., 24, 4256-4262 (1996)), 5' of its puromycin resistance gene and loxP site, giving rise to plasmid pRK52. The sequence and orientation of the cloned attB site was confirmed by DNA sequence analysis. Next, an attP site site (Thorpe et al. Proc. Natl. Acad. Sci. USA, 95, 5505 - 5510 (1998)), generated by the annealing of the two synthetic oligonucleotides C31-6 (SEQ ID NO:3) and C31-7-2 (SEQ ID NO:4), was ligated into the BamHI restriction site of plasmid pRK52, downstream of the puromycin resistance gene and loxP site, giving rise to plasmid pRK64 (SEQ ID NO:5). The sequence and orientation of the attP site was confirmed by DNA sequence analysis. In pRK64 the newly cloned attB (position 348 - 431) and attP (position 1534 - 1617) sites are in the same orientation flanking the puromycin resistance gene and the SV40 early polyadenylation sequence. The attB and attP sites are followed by loxP sites in the same orientation (positions 435 - 469 and 1624 - 1658). The puromycin cassette is transcribed from the SV40 early enhancer/promoter region and followed by the coding region for E. coli B-galactosidase and the SV40 late region polyadenylation sequence.

Construction of C31-Int expression vectors: First the C31-Int gene of phage ΦC31 was amplified by PCR from phage DNA (DSM-49156, DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-38124 Braunschweig, Germany) using the primers C31-1 (SEQ ID NO:6) and C31-3 (SEQ ID NO:7). The ends of the PCR product were digested with NotI and the product was ligated into plasmid pBluescript II KS, opened with NotI, giving rise to plasmid pRK40. The DNA sequence of the 1.85 kb insert was determined and found to be identical to the published C31-Int gene (Kuhstoss et al., J. Mol. Biol. 222, 897-908 (1991)), except for an error in the stop codon. This error was repaired by PCR amplification of a 300 bp fragment from plasmid pRK40 using the primers C31-8 (SEQ ID NO:8) and C31-9 (SEQ ID NO:9), which provide a

corrected Stop codon. The ends of this PCR fragment were digested with Eco47III and XhoI, the fragment was ligated into plasmid pRK40 and opened with Eco47III and XhoI to remove the fragment containing the defective stop codon. The resulting plasmid pRK55 contains the correct C31-Int gene as confirmed by DNA sequence analysis.

5

10

30

35

The C31-Int gene was isolated from pRK55 as 1.85 kb fragment by digestion with NotI and XhoI and ligated into the generic mammalian expression vector pRK50 (see below), opened with NotI and XhoI, giving rise to the C31-Int expression vector pCMV-C31-Int(wt). pCMV-C31-Int(wt) (SEQ ID NO:10) contains a 700 bp cytomegalovirus immediated early gene promoter (position 1-700), a 270 bp hybrid intron (position 701-970), the C31-Int gene (position 978-2819), and a 189 bp synthetic polyadenylation sequence (position 2831-3020).

For the construction of pCMV-C31-Int(NNLS) a 1.5 kb fragment was amplified by PCR from phage DNA using oligonucleotides C31-2 (SEQ ID NO:98) and C31-3 (SEQ ID NO:7). The ends of the PCR product were digested with NotI and the product was ligated into plasmid pBluescript II KS, opened with NotI, giving rise to plasmid pRK41 (SEQ ID NO: 99). A 1100 bp fragment generated by digestion of plasmid pRK41 with NcoI and NotI was then ligated into plasmid pRK55 (SEQ ID NO:80), opened with NcoI and NotI, giving rise to the plasmid pRK63 (SEQ ID NO:81). The C31-Int gene with N-terminal NLS was isolated as a 1.8 kb fragment from pRK63 by digestion with NotI and XhoI and ligated into the mammalian expression vector pRK50, opened with NotI and XhoI, giving rise to the C31-Int expression vector pCMV-C31-Int(NNLS). pCMV-C31-Int(NNLS) (SEQ ID NO:73) contains a 700 bp Cytomegalovirus immediated early gene promoter (position 1 – 700), a 270 bp hybrid intron (position 701 – 970), the C31-Int gene with N-terminal NLS (position 976 – 2838), and a 189 bp synthetic polyadenylation sequence (position 2851 – 3040).

For the construction of pCMV-C31-Int(CNLS), the 3'-end of the C31-Int gene was amplified from pCMV-C31-Int(wt) as a 300 bp PCR fragment using the primers C31-8 (SEQ ID NO:8) and C31-2-2 (SEQ ID NO:11). Primer C31-2-2 modifies the 3'-end of the wildtype C31-Int gene such that the stop codon is replaced by a sequence of 21 basepairs coding for the SV40 T-antigen nuclear localisation sequence of 7 amino acids (Prolin-Lysin-Lysin-Lysin-Arginin-Lysin-Valin) (Kalderon et. al, Cell, 39, 499 – 509 (1984)), followed by a new stop

codon. The ends of this 300 bp PCR fragment were digested with with Eco47III and XhoI, the fragment was ligated into plasmid pCMV-C31-Int(wt) and opened with Eco47III and XhoI to replace the 3'-end of the wildtype C31-Int gene resulting in the plasmid pCMV-C31-Int(CNLS). The identity of the new gene segment was verified by DNA sequence analysis. pCMV-C31-Int(CNLS) (SEQ ID NO:12) contains a 700 bp cytomegalovirus immediated early gene promoter (position 12 – 711), a 270 bp hybrid intron (position 712 – 981), the modified C31-Int gene (position 989 – 2851), and a 189 bp synthetic polyadenylation sequence (position 2854 – 3043).

10

15

20

25

5

To generate the Cre expression plasmid pCMV-Cre (SEQ ID NO:13), the coding sequence of Cre recombinase (Sternberg et al., J. Mol. Biol., 187, 197 - 212 (1986)) with a N-terminal fusion of the 7 amino acid SV40 T-antigen NLS (see above) was recovered from plasmid pgk-Cre and cloned into the NotI and XhoI sites of plasmid pRK50. PRK50 (SEQ ID NO:14) is a generic expression vector for mammalian cells based on the cloning vector pNEB193 (New England Biolabs Inc, Beverly, MA, USA). PRK50 was built by insertion into pNEB193 of a 700 bp cytomegalovirus immediated early gene (CMV-IE) promoter (position 1-700) from plasmid pIREShyg (GenBank#U89672; Clontech Laboratories Inc, Palo Alto, CA, USA), a synthetic 270 bp hybrid intron (position 701-970), consisting of a adenovirus derived splice donor and an IgG derived splice acceptor sequence (Choi et al., Mol. Cell. Biol., 11, 3070 - 3074 (1991)), and a 189 bp synthetic polyadenylation sequence (position 1000-1188) build from the polyadenylation consensus sequence and 4 MAZ polymerase pause sites (Levitt et al., Genes&Dev., 3, 1019 - 1025 (1989); The EMBO J. 13, 5656 - 5667 (1994)). The positive control plasmid pRK64(Δ Cre) (SEQ ID NO:15) was generated from pRK64 by transformation into the Cre expressing E. coli strain 294-Cre (Buchholz et al., Nucleic Acids Res., 24, 3118 - 3119 (1996)).

One of the transformed subclones was confirmed for the Cre mediated deletion of the loxP-flanked cassette by restriction mapping and further expanded. Plasmid pUC19 is a cloning vector without eukaryotic control elements used to equalise DNA amounts for transfections (GenBank#X02514; New England Biolabs Inc, Beverly, MA, USA). All plasmids were propagated in DH5α E. coli cells (Life Technologies GmbH, Karlsruhe, Germany) grown in Luria-Bertani medium and

WO 02/38613 PCT/EP01/12975

purified with the plasmid DNA purification reagents "Plasmid-Maxi-Kit" (Quiagen GmbH, Hilden, Germany) or "Concert high purity plasmid purification system" (Life Technologies GmbH, Karlsruhe, Germany). Following purification, the plasmid DNA concentrations were determined by absorption at 260 nm and 280 nm in UVette cuvettes (Eppendorf-Netheler-Hinz GmbH, Hamburg, Germany) using a BioPhotometer (Eppendorf-Netheler-Hinz GmbH, Hamburg, Germany) and the plasmids were diluted to the same concentration; finally these were confirmed by separation of 10 ng of each plasmid on an ethidiumbromide-stained agarose gel.

10

20

25

30

35

B. Cell culture and transfections: Chinese hamster ovary (CHO) cells (Puck et al., J. Exp. Med., 108, 945 (1958)) were obtained from the Institute for Genetics (University of Cologne, Germany) as a population adapted to growth in DMEM medium. The cells were grown in DMEM/Glutamax medium (Life Technologies) 15 supplemented with 10% fetal calf serum at 37°C, 10% CO2 in humid atmosphere and passaged upon trypsinisation. One day before transfection 10⁶ cells were plated into a 48-well plate (Falcon). For the transfection of cells with plasmids each well received into 250 ml of medium a total amount of 300 ng supercoiled plasmid DNA complexed before with the FuGene6 transfection reagent (Roche Diagnostics GmbH, Mannheim, Germany) according to the manufacturers protocol. Each 300 ng DNA preparation (Fig.2 sample 4 to 11) contained 50 ng of the luciferase expression vector pUHC13-1 (Gossen et al., Proc Natl Acad Sci USA., 89 5547-5551 (1992)), 50 ng of the substrate vector pRK64, 0.5 ng or 1 ng of one of the recombinase expression vectors pCMV-C31Int(wt), pCMV-C31Int(NNLS), pCMV-C31Int(CNLS) or pCMV-Cre and 199 ng or 199.5 ng of pUC19 plasmid, except for the controls which received 50 ng of pUHC13-1 together with 50 ng of pRK64 (sample 3) or pRK64(Δcre) (sample 2) and 200 ng pUC19, or 50 ng pUHC13-1 with 250 ng pUC19 (sample 1). Transfections of Ssv and XisA recombinases (Fig.3) also contained 50 ng of the luciferase expression vector pUHC13-1, 50 ng of substrate vectors pPGKattA and pPGKnif and 10 ng or 20 ng of recombinase expression vector pCMV-SSV or pCMV-SSV(NNLS) or 25 ng or 100 ng of expression vectors pCMV-XisA/ pCMV-XisA(NNLS). Plasmid pUC19 was added to a total amount of 300 ng plasmid DNA. As the C31-Int expression vectors are 15% larger in size than pCMV-Cre and the same amounts of DNA of the three plasmids were used for transfection, the

samples with C31-Int vectors received 15% less plasmid molecules as compared to the samples with Cre expression vector. The ß-galactosidase values from C31-Int transfected samples by this value were not corrected and thus is a slight underestimation of the calculated C31-Int activities. For each sample to be tested four individual wells were transfected. One day after the addition of the DNA preparations each well received additional 250 ml of growth medium. The cells of each well were lysed 48 hours after transfection with 100 ml lysate reagent supplemented with protease inhibitors (Roche Diagnostics). The lysates were centrifuged and 20 ml were used to determine the ß-galactosidase activities using the β -galactosidase reporter gene assay (Roche Diagnostics) according to the manufacturers protocol in a Lumat LB 9507 luminometer (Berthold). To measure luciferase activity, 20ml lysate was diluted into 250ml assay buffer (50mM glycylglycin, 5mM MgCl₂, 5mM ATP) and the "Relative Light Units" (RLU) were counted in a Lumat LB 9507 luminometer after addition of 100 ml of a 1 mM luciferin (Roche Diagnostics) solution. The mean value and standard deviation of the samples was calculated from the ß-galactosidase and luciferase RLU values obtained from the four transfected wells of each sample.

10

C. Results: To set up an assay system for the measurement of C31-Int and Cre recombinase efficiency in mammalian cells the recombination substrate vector 20 pRK64 shown in Figure 1E was first constructed. pRK64 contains a SV40 promoter region for expression in mammalian cells followed by a 1.1 kb cassette which consists of the coding region of the puromycin resistance gene and a polyadenylation signal sequence. This cassette is flanked at the 5`-end by the 84 bp attB and at the 3'-end by the 84 bp attP recognition site of C31-Int (Fig.1 and 6). These attB and attP sites are located on the same DNA molecule and oriented in a way to each other which allows the deletion of the flanked DNA segment. The same orientation of attB and attP sites is used naturally by the Φ C31 phage and the bacterial genome, leading to the integration of the phage genome when both sites are located on different DNA molecules (Thorpe et al., 30 Proc. Natl. Acad. Sci. USA, 95, 5505 - 5510 (1998)). To measure C31-Int and Cre recombinase activities with the same substrate vector, pRK64 contains in addition two Cre recognition (loxP) sites in direct orientation next to the att sites. Since the att/lox-flanked cassette in plasmid pRK64 is inserted between the SV40 promoter and the coding region of the ß-galactosidase gene, its presence inhibits 35

30

35

 β -galactosidase expression as the SV40 promoter derived transcripts are terminated at the polyadenylation signal of the puromycin gene. Plasmid pRK64 is turned into a β -galactosidase expression vector upon C31-Int or Cre mediated deletion of the att/lox-flanked puromycin cassette since the remaining single att and loxP site do not substantially interfere with gene expression.

28

For the expression of recombinases a mammalian expression vector was designed which contains the CMV immediate early promoter followed by a hybrid intron, the coding region of the recombinase to be tested, and an artificial polyadenylation signal sequence. The backbone sequence of the four 10 recombinase expression vectors shown in Figure 1A-D is identical to each other except for the recombinase coding region. Plasmid pCMV-C31Int(wt) (Fig. 1A) contains the nonmodified (wildtype) 1.85 kb coding region of C31-Int as found in the genome of phage Φ C31 (Kuhstoss, et al., J. Mol. Biol. 222, 897-908 (1991)). Plasmid pCMV-C31Int(NNLS) (Fig. 1B) contains a modified C31-Int gene coding for the full length C31-Int protein with a N-terminal extension of 7 amino acids derived from the SV40 virus large T antigen which serves as a nuclear localisation signal (NLS). Plasmid pCMV-C31Int(CNLS) (Fig. 1C) contains a Cterminal extension of 7 amino acids derived from the SV40 virus large T antigen which serves as a nuclear localisation signal (NLS). Plasmid pCMV-Cre (Fig. 1D) 20 contains the 1.1 kb Cre coding region with an N-terminal fusion of the 7 amino acid NLS of the SV40 T-antigen. For Cre recombinase it has been shown that the N-terminal addition of the SV40 T-antigen NLS does not increase its recombination efficiency in mammalian cells (Le et al., Nucleic Acids Res., 27, 25 4703 - 4709 (1999)).

As a test system to compare the efficiency of the 4 recombinases the same amount of plasmid DNA of each of the recombinase expression vectors together with a fixed amount of the reporter plasmid pRK64 was transiently introduced into Chinese Hamster Ovary (CHO) cells. Thus, in this assay design the efficiency of the various recombinases on an extrachromosomal substrate introduced into the CHO cells was compared as a circular plasmid. Two days after transfection the cells from the various samples were lysed and the activity of β -galactosidase in the lysates was determined by a specific chemiluminescense assay and expressed in "Relative Light Units" (RLU (β -Gal)) (Fig. 2). In addition all samples

contained a fixed amount of a luciferase expression vector to control for the experimental variation of cell transfection and lysis. For this purpose the lysates of each sample were also tested for luciferase activity with a specific chemiluminescense assay and the values expressed as "Relative Light Units" (RLU (Luciferase)) (Fig. 2). All transfection samples contained in addition varying amounts of the unrelated cloning plasmid pUC19 so that all samples were equalised to the same amount of plasmid DNA. As a positive control for β -galactosidase a derivative of the recombination reporter pRK64 was used in which the loxP flanked 1.1 kb cassette has been removed through Cre mediated recombination in E. coli giving rise to plasmid pRK64(Δ Cre). As negative controls served samples which received the unrecombined reporter plasmid pRK64 but no recombinase expression vector as well as samples set up with the pUC19 plasmid alone.

10

25

30

35

To determine the relative efficiency of the tested recombinases the RLU values of β-galactosidase were divided individually for each sample by the RLU values obtained for luciferase and multiplied with 10⁵. From the values of the four data points of each sample the mean value and standard deviation was calculated as an indicator of recombinase activity (Gal/Luc) (Fig. 2). The relative activity of the tested recombinases was then compared to the positive control defined as an activity of 1.

As shown in Fig. 2, the expression of Cre recombinase (samples 10 and 11) resulted in a 150 to 170-fold increase of β -galactosidase activity as compared to the negative control (sample 3), demonstrating the wide dynamic range of our test system. Each recombinase vector was tested using two different amounts of DNA for transfection (0.5 and 1ng/sample), which in the case of Cre resulted in 63% and 72% recombinase activity (samples 10 and 11 as compared to the positive control). These two values establish that the DNA amounts used are close to the test systems saturation for recombinase expression as the doubling of DNA amounts resulted only in a minor increase of recombinase activity.

In comparison to Cre, the expression of wildtype C31-Int resulted in a considerably lower recombinase activity of 23% and 30% (Fig. 2, samples 4 and 5) as compared to the positive control. These values represent 37% and 42%

recombinase activity for wildtype C31-Int as compared to Cre recombinase (compare samples 4 and 5 with 10 and 11). Upon the expression of C31-Int fused with the N-terminal NLS (C31-Int(NNLS)) values of 32% and 36% recombinase activity (samples 6 and 7) were obtained as compared to the positive control. The C31-Int(NNLS) values represent 51% and 50% recombinase activity as compared to Cre (compare samples 6 and 7 to 10 and 11). Thus, the activity of C31-Int in mammalian cells is just moderately enhanced by the addition of a NLS signal.

Surprisingly, upon the expression of C31-Int fused with the C-terminal NLS (C31-Int(CNLS)) values of 50% and 65% recombinase activity (samples 8 and 9) were obtained as compared to the positive control. The C31-Int(CNLS) values represent 79% and 90% recombinase activity as compared to Cre recombinase (compare samples 84 and 9 to 10 and 11). Unexpectedly,C31-Int(CNLS) exhibits a dramatic, more than twofold increase of recombinase activity in comparison to C31-Int(wt) (compare samples 8 and 9 to 4 and 5).

10

15

20

25

30

35

In order to test whether the addition of a NLS sequence may be a general, simple method to enhance recombinase activity in mammalian cells we extended our studies by two additional recombinases: XisA recombinase (XisA) derived from the cyanobacterium Anabaena, and SSV-Integrase (SSV-Int) derived from the SSV1 virus of the thermophilic bacterium Sulfolobus shibatae. To this end we constructed mammalian expression vectors for the wildtype forms of XisA and SSV recombinases and compared their activity to versions which were modified by the N-terminal addition of the 7 amino acid NLS of the SV40 T-antigen. These recombinases were compared by the use of the reporter vector shown in Fig.1E, except that the att elements of C31-Int were replaced by the nif recognition sequences for XisA or the att sequences for SSV-Int. As described above for C31-Int, recombinase activities were tested by transient transfection into CHO cells using the reporter vector derived β -galactosidase activity as readout and cotransfected luciferase as internal control.

As shown in Fig.3 for both, XisA and SSV recombinases the addition of a NLS sequence did not improve their activity in a mammalian cell line as compared to the wildtype forms. At both DNA concentrations tested wildtype XisA exhibits a significant recombination activity as compared to the reporter vector alone (compare samples 2 and 3 to sample 1), but this activity is not altered by the

addition of an NLS (compare samples 2 and 3 to samples 4 and 5). SSV-Int exhibits only a low recombination activity (compare samples 7 and 8 with sample 6) which is also not enhanced by the addition of a NLS (compare samples 9 and 10 with samples 7 and 8). From these results we conclude that the addition of a NLS to an inefficient recombinase is not a general, simple method to improve its performance in mammalian cells.

Taken together, in the transient transfection test system shown in Figure 2 a more than twofold activity increase of the Φ C31 Integrase could be achieved by the C-terminal, but not the N-terminal addition of the SV40 T antigen NLS signal. As this signal sequence has been characterised to act as a nuclear localisation signal (Kalderon et. al, Cell, 39, 499 – 509 (1984)) we conclude that the efficiency increase of C31-Int(CNLS) is the result of the improved nuclear accumulation of this recombinase. The relative inefficiency of C31-Int (NNLS) may be explained by the inaccessibility of the NLS peptide to the nuclear import machinery at the N-terminal position of the C31-Int protein.

In particular, it could be shown that C31-Int(CNLS) recombines extrachromosomal DNA in mammalian cells almost as efficient as the widely used Cre recombinase and thus provides an additional or alternative recombination system of highest activity. The efficiency increase of C31-Int(CNLS) as compared to its wildtype form is regarded as an invention of substantial use for biotechnology.

Example 2

5

10

15

20

As demonstrated in example 1 C31-Int recombinase with the C-terminal fusion of the SV40 T-antigen NLS (C31-Int(CNLS)) shows in mammalian cells a recombination activity comparable to Cre recombinase on an extrachromosomal plasmid vector. It was further tried to test whether C31-Int(CNLS) exhibits a similar activity on a recombination substrate which is chromosomally integrated into the genome of mammalian cells. This question is critical for the use of a recombination system for genome engineering as it is possible that a recombinase may act efficiently on extrachromosomal substrates but is impaired if the recognition sites are part of the mammalian chromatin. To characterise the recombination activity of C31-Int(CNLS) and C31-Int(NNLS) on a chromosomal substrate the pRK64 reporter plasmid (Fig. 1E) was stably integrated, containing

WO 02/38613

a pair of loxP and att sites, into the genome of a mammalian cell line. One of the stable transfected clones was chosen for further analysis and was transiently transfected with recombinase expression vectors coding for C31-Int(CNLS), C31-Int(NNLS), C31-Int(wt) or Cre recombinase. The activity of ß-galactosidase derived from the Cre expression vector recombined in these cells was taken as a measure of recombination efficiency.

A. Plasmid constructions: all plasmids used and their purification are described in example 1.

10

25

30

35

5

B. Cell culture and transfections: To generate a stably transfected C31-Int reporter cell line 2.5×10^6 NIH-3T3 cells (Andersson et al., Cell, 16, 63-75(1979); DSMZ#ACC59; DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-38124 Braunschweig, Germany) were electroporated with 5 µg pRK64 plasmid DNA linearised with ScaI and plated into 10cm petri dishes. The cells were grown in DMEM/Glutamax medium (Life Technologies) supplemented with 10% fetal calf serum at 37°C, 10% CO_2 in humid atmosphere, and passaged upon trypsinisation. Two days after tranfection the medium was supplemented with 1mg/ml of puromycin (Calbiochem) for the 20 selection of stable integrants. Upon the growth of resistant colonies these were isolated under a stereomicroscope and individually expanded in the absence of puromycin. To demonstrate stable integration of the transfected vector, genomic DNA of puromycin resistant clones was prepared according to standard methods and $5-10~\mu g$ were digested with EcoRV. Digested DNA was separated in a 0.8%agarose gel and transferred to nylon membranes (GeneScreen Plus, NEN DuPont) under alkaline conditions for 16 hours. The filter was dried and hybridised for 16 hours at 65°C with a probe representing the 5` part of the E. coli ß-galactosidase gene (1.25 kb NotI - EcoRV fragment of plasmid CMV-ß-pA (R. Kühn, unpublished). The probe was radiolabelled with P32-marked $\alpha\text{-dCTP}$ (Amersham) using the Megaprime Kit (Amersham). Hybridisation was performed in a buffer consisting of 10% dextranesulfate, 1% SDS, 50 mM Tris and 100 mM NaCl, pH7.5). After hybridisation the filter was washed with 2x SSC/1%SDS and exposed to BioMax MS1 X-ray films (Kodak) at ~ 80°C.

Transfections of the selected clone 3T3(pRK64)-3 with plasmid DNAs and the measurement of B-galactosidase activities in lysates were essentially performed WO 02/38613

as described in example 1 for CHO cells, except that 32ng or 64ng of the recombinase expression plasmids and 218 or 186 ng of pUC19 plasmid were used and the pRK64 plasmid was omitted from all samples.

33

5 <u>C. Histochemical detection of β-galactosidase activity in transfected 3T3(pRK64)-3 cells</u>

To directly demonstrate β -galactosidase expression in recombinase transfected cells, 10^4 3T3(pRK64)-3 cells were plated one day before transfection into each well of a 48-well tissue culture plate (Falcon). For the transient transfection of cells with plasmids each well received into 250 µl of medium a total amount of 150 ng supercoiled plasmid DNA complexed before with the FuGene6 transfection reagent (Roche Diagnostics GmbH, Mannheim, Germany) according to the manufacturers protocol. Each 150 ng DNA preparation contained 50 ng of the recombinase expression vector pCMV-Cre or pCMV-C31Int(CNLS) and 100ng of the pUC19 plasmid. After 2 days the culture medium was removed from the wells, the wells were washed once with phosphate buffered saline (PBS), and the cells were fixed for 5 minutes at room temperature in a solution of 2% formaldehyde and 1% glutaraldehyde in PBS. Next, the cells were washed twice with PBS and finally incubated in X-Gal staining solution for 24 hours at 37°C (staining solution: 5 mM K_3 (Fe(CN)₆), 5 mM K_4 (Fe(CN)₆), 2 mM MgCl₂, 1mg/ml X-Gal (BioMol) in PBS) until photographs were taken.

D. Results

10

15

20

To generate a mammalian cell clone with a stable genomic integration of the C31-Int and Cre recombinase reporter plasmid pRK64, the murine fibroblast cell line NIH-3T3 was electroporated with linearised pRK64 DNA (Fig.1D; see also example 1) and subjected to selection in puromycin containing growth medium. Plasmid pRK64 contains in between the pair of loxP and att sites the coding region of the puromycin resistance gene expressed from the SV40-IE promoter.

Thirty-six puromycin resistant clones were isolated and the genomic DNA of 19 clones was analysed for the presence and copy number of the pRK64 DNA. Three clones, which apparently contain 2 - 4 copies of pRK64, were further characterised on the single cell level for the expression of β-galactosidase upon transient transfection with the Cre expression vector pCMV-Cre (Fig. 1C). The cell clone with the largest proportion of β-galactosidase positive cells, 3T3(pRK64)-3,

was selected as most useful for the planned studies on C31-Int and Cre recombinase efficiency.

To compare the efficiency of wildtype C31-Int (C31-Int(wt)), C31-Int(CNLS), C31-Int(NNLS), and Cre recombinases 32ng or 64 ng of the recombinase expression vectors pCMV-C31Int(wt), pCMV-C31Int(CNLS), pCMV-C31Int(NNLS), or pCMV-Cre (Fig. 1 A-D) together with the unrelated cloning plasmid pUC19 were transiently introduced into 3T3(pRK64)-3 cells, such that all samples contained the same amount of plasmid DNA. As a negative control a sample prepared with the pUC19 plasmid alone was used. Two days after transfection the cells from the various samples were lysed and the activity of β-galactosidase in the lysates was determined by a specific chemiluminescense assay and expressed in "Relative Light Units" (RLU)(β-Gal) (Fig. 4). From the values of the four data points of each sample the mean value and standard deviation was calculated as an indicator of recombinase activity (Fig.4). The relative activity of the tested recombinases was then compared to the highest value obtained with the Cre expression vector, defined as an activity of 1.

10

15

20

25

30

35

As shown in Figure 4 the expression of Cre recombinase (samples 8 and 9) resulted in a 36 to 49-fold increase of β-galactosidase activity as compared to the negative control (sample 1), demonstrating the dynamic range of the test system used. Each recombinase vector was tested using two different amounts of DNA for transfection (32 ng and 64 ng/sample), which in the case of Cre resulted in 73% and 100% recombinase activity (samples 8 and 9). These two values establish that the DNA amounts used are not far from the linear scale of the test systems ability for recombinase expression as the twofold increase of the amount of DNA also resulted in a significant increase of recombinase activity.

The expression of wildtype C31-Int (Fig. 4, samples 2 and 3) resulted in a low recombinase activity of 4% and 10% as compared to the values obtained by Cre transfection (compare samples 2 and 3 with 8 and 9). This activity was only moderately enhanced by the expression of C31-Int(NNLS) to values of 19% and 22% of Cre activity (compare samples 4 and 5 with samples 8 and 9). Upon the expression of C31-Int(CNLS) values of 48% and 78% recombinase activity were obtained as compared to Cre recombinase (compare samples 6 and 7 to 8 and

9). Hence, C31-Int(CNLS) exhibits an 12-fold higher activity than C31-Int(wt) at 32 ng plasmid DNA (Fig.4, compare samples 6 and 2) and an 8-fold higher activity than C31-Int(wt) at 64 ng plasmid DNA (Fig.4, compare samples 7 and 3).

5

10

20

25

In addition, it was aimed to directly demonstrate in situ the expression of ßgalactosidase in 3T3(pRK64)-3 cells after transfection with Cre or C31-Int(CNLS) recombinase plasmid. Two days after transfection the cells were fixed in situ and incubated with the histochemical X-Gal assay which detects β-galactosidase positive cells by a blue precipitate. As shown in Figure 5 stained cells were found at a comparable frequency in the samples transfected with the Cre or C31-Int(CNLS) expression vectors but not in the nontransfected control. This result confirms that the B-galactosidase activities measured by chemiluminescense upon recombinase transfection (Fig. 4) results from a population of individual, recombined reporter cells.

- 15

In conclusion, upon the transient transfection of recombinase expression vectors into a cell line with a genomic integration of the recombination substrate vector, a 8 – 12-fold activity increase of the Φ C31 Integrase by the C-terminal fusion with the SV40 T-antigen NLS signal was found. As this signal sequence has been characterised to act as a nuclear localisation signal (Kalderon et. al, Cell, 39, 499 - 509 (1984)), it was concluded that the dramatic efficiency increase of C31-Int(CNLS) is the result of the improved nuclear accumulation of this recombinase. The approximately tenfold activity increase of C31-Int(CNLS) upon expression in a cell line with a genomic integration of the substrate vector is considerably higher than the activity increase found upon the transient expression of both vectors (see example 1). Thus, a substrate vector integrated into the chromatin of a mammalian cell may pose more stringent requirements on recombinase activity to be recombined as compared to an extrachromosomal substrate.

30

The dramatic activity increase of C31-Int(CNLS), as compared to its wildtype form, on a stable integrated substrate in mammalian cells is an invention of significant practical use as this recombinase is as efficient as the widely used

36

Cre/loxP system; thus, C31-Int(CNLS) provides an additional or alternative recombination system of highest activity.

Example 3

To demonstrate that the increase in β-galactosidase activity obtained by the cotransfection of a C31-Int expression vector and the reporter vector pRK64 into mammalian cells is in fact the result of recombinase mediated deletion, one of the recombination products was detected by a specific polymerase chain reaction (PCR). The amplified PCR product was cloned and its sequence determined. The obtained sequence confirms that C31-Int mediated deletion of the test vector occurs in a mammalian cell line and that the recombination occurs at the known breakpoint within the attB and attP sites.

A. Plasmid constructions: The construction of plasmids pRK64, pCMV-Cre and pCMV-C31-Int(wt) is described in Example 1. To simulate the recombination of pRK64 by C31-Int, the sequence between the CAA motives of the att sites (boxed in Fig.5) was deleted from the computerfile of pRK64, giving rise to the sequence of pRK64(ΔInt) (SEQ ID NO:16).

B. Transfection of Cells and PCR amplification: MEF5-5 mouse fibroblasts 20 (Schwenk et al., 1998) (20000 cells per well of a 12 well plate (Falcon)) were transfected with 0.5 μg pRK64 alone or together with 250 ng pCMV-Int(wt) or pCMV-Cre using the FuGene6 transfection reagent following the manufacturers protocol (Roche Diagnostics). After 2 days DNA was extracted from these cells according to standard methods and used for PCR amplification with Primers P64-25 1 (SEQ ID NO:17; complementary to position 111-135 of pRK64(Δ Int)) and P64-4 (SEQ ID NO:18; complementary to position 740-714 of pRK64(Δ Int)) using the Expand High Fidelity PCR kit (Roche Diagnostics). PCR products were separated on a 0.8% agarose gel, extracted with the QuiaEx kit (Quiagen) and cloned into the pCR2.1 vector using the TA cloning kit (Invitrogen) resulting in plasmid 30 pRK80d. The sequence of its insert, seq80d (SEQ ID NO:19), was determined using the reverse sequencing primer and standard sequencing methods (MWG Biotech).

35 For the measurement of β-galactosidase activity the cells were lysed 2 days after

transfection and the β -galactosidase activities were determined with the β -galactosidase reporter gene assay (Roche Diagnostics) as described in example 1.

C. Results: As a test vector for C31-Int mediated DNA recombination plasmid pRK64 was used, which contains the 1.1 kb coding region of the puromycin resistance gene flanked 5' by the 84 bp attB and 3' by the 84 bp attP recognition site of C31-Int (Fig. 5). These attB and attP sites are located on the same DNA molecule and oriented in a way to each other which allows the deletion of the att-flanked DNA segment. The same orientation of attB and attP 10 sites is used naturally by the Φ C31 phage and the bacterial genome for the integration of the phage genome when both sites are located on different DNA molecules (Thorpe et al., Proc. Natl. Acad. Sci. USA, 95, 5505 - 5510 (1998)). As a positive control, vector pRK64 contains in addition two Cre recombinase recognition (loxP) sites in direct orientation next to the att sites. Since the att-15 flanked DNA segment in plasmid pRK64 is inserted between a promoter active in mammalian cells and the β -galactosidase gene, its deletion can be measured by the increase of B-galactosidase activity. The expected product of C31-Int mediated deletion of plasmid pRK64 is shown in Fig. 6, designated as pRK64(ΔInt). If the recombination between attB and attP occurs as described in 20 bacteria (Thorpe et al., Proc. Natl. Acad. Sci. USA, 95, 5505 - 5510 (1998)), a single attR site is generated and left on the parental plasmid (Fig. 6) while the flanked DNA is excised and contains an attL site. Beside the measurement of Bgalactosidase activity, C31-Int mediated recombination of pRK64 can be directly detected on the DNA level by a specific polymerase chain reaction (PCR) using the primers P64-1 and P64-4 (Fig. 6). These primers, located 5 of the attB site (P64-1) and 3' of the attP site, are designed to amplify a PCR product of 630 bp lenght upon the C31-Int mediated recombination of pRK64. For the expression of C31-Int in mammalian cells plasmid pCMV-C31(wt) was used, which contains the 30 CMV-IE-Promoter upstream of the C31-Int coding region followed by a synthetic polyadenylation signal (see Example 1 and Fig.1).

The recombination substrate vector pRK64 was transiently transfected into the murine fibroblast cell line MEF5-5 either alone, or together with the C31-Int expression vector pCMV-C31(wt), or together with an expression vector for Cre recombinase, pCMV-Cre. Two days after transfection half the cells of each sample

WO 02/38613

5

15

20

35

was lysed and used to measure ß-galactosidase activity by chemiluminescense, and the other half was used for the preparation of DNA from the transfected cells for PCR analysis. The ß-galactosidase levels of the 3 samples were found as following (expressed as "Relative Light Units" (RLU) with standard deviation (SD) of the ß-galactosidase assay):

| | Sample | RLU (SD) |
|----|-------------------------|-------------------|
| | 1) pRK64 | 692 <u>+</u> 5 |
| | 2) pRK64 + pCMV-Cre | 8527 <u>+</u> 269 |
| 10 | 3) pRK64 + pCMV-C31(wt) | 1288 <u>+</u> 93 |

As the coexpression of the test vector pRK64 together with the C31-Int expression vector in sample 3 leads to a significant increase of ß-galactosidase activity as compared to pRK64 alone, this result suggests that pRK64 is recombined by C31-Int as anticipated in Fig. 6.

Next, cellular DNA was prepared from the three samples and tested for the occurrence of the expected Cre or C31-Int generated deletion product by PCR using primers P64-1 and P64-4 for amplification. As shown in Fig. 7 an amplification product of the expected size was found only in the samples cotransfected with the Cre or C31-Int recombinase expression vectors (Fig. 7A, lane3 and lane 4). The PCR products amplified from pRK64 recombined by C31-Int or Cre are of the same size but should be recombined via the attB/P or loxP sites, respectively.

To prove that the PCR product found after cotransfection of plasmids pRK64 and pCMV-C31(wt) represents in fact the deletion product of C31-Int mediated recombination, this DNA fragment was cloned into a plasmid vector and its DNA sequence determined. One clone, pRK80d, was analysed, and its sequence showed exactly the sequence of an attR site as expected from C31-Int mediated deletion of pRK64 (Fig. 7B, compare to Fig. 6).

In conclusion, this experiment demonstrates that C31-Int mediated deletion of a vector containing a pair of attB/attP sites occurs in a mammalian cell line, and that the recombination occurs within the same 3 bp breakpoint region of attB and attP as found in bacteria (Thorpe et al., Proc. Natl. Acad. Sci. USA, 95, 5505 ~

5510 (1998)). Thus, it was concluded that an increase of β -galactosidase activity observed by cotransfection of the pRK64 reporter vector and a C31-Int expression vector in mammalian cells truly reflects C31-Int recombinase activity.

5

10

15

30

35

Example 4

As has been demonstrated in examples 1-3, the C31-Int recombinase with the C-terminal fusion of the SV40 T-antigen NLS (C31-Int(CNLS)) shows a recombination activity comparable to Cre recombinase on an extrachromosomal as well as a chromosomally integrated target in mammalian cells in vitro. To test whether C31-Int(CNLS) exhibits activity in mice, transgenic mice carrying a C31-Int(CNLS) expression vector were generated. These transgenic mice were crossed with reporter mice carrying the recombinase substrate. Recombination-mediated expression of β -galactosidase, which can be measured by staining with the substrate X-Gal, was analyzed in testes of double transgenic progeny carrying both the recombinase and the reporter.

A. Plasmid constructions: For the construction of the C31-Int(CNLS) transgene expression vector, the C31Int gene with C-terminal NLS was isolated as a 2 kb-fragment generated by restriction of pCMV-C31Int(CNLS) (SEQ ID NO: 12) with BgIII. The fragment was ligated into the BgIII restriction site of the vector pCAGGS-Cre-pA (SEQ ID NO:104) giving rise to the plasmid pCAGGS-C31CNLS-pA (SEQ ID NO:105). In pCAGGS-C31CNLS-pA the C31-Int(CNLS) (position 1891-3753) is transcribed from the CAGGS promoter (position 1-1616) and followed by the SV40 late region polyadenylation sequence (position 3763-3941).

B. Production of transgenic mice: For the embryo injections a 3.95 kb-fragment was generated by restriction of the plasmid pCAGGS-C31CNLS-pA with PstI and AscI. This fragment was purified as follows: DNA bands were separated on an agarose-gel without ethidiumbromide. One part of the gel was stained with ethidiumbromide to locate the band to excise. The DNA was electroeluted from the excised band with S&S Biotrap Elution Chamber in 1x TAE (40 mM Trisacetate, 1 mM EDTA) overnight. The DNA was precipitated from the eluate with 1/10 volume 3M sodium acetate and 2.5 volumes ethanol at -20 °C for several

hours. The DNA was pelleted by centrifugation at 13000 rpm for 30 min and washed twice with 70 % ethanol. The dried DNA pellet was resuspended in TE (10 mM Tris, 1 mM EDTA, pH 8). Subsequently the precipitation procedure was repeated once and the DNA resuspended in injection buffer (10 mM Tris pH 7.2, 0.1 mM EDTA). The sample was dialysed with Slide-A-Lyse Mini Dialysis Unit (Pierce) in injection buffer with several changes of buffer at 4°C overnight. Different amounts of the sample were checked on a gel to determine concentration. To generate transgenic mice, 5-10 fg of the purified fragment were injected into one pronucleus of (B6CBA)F2 mouse one-cell embryos. The injected embryos were subsequently transferred into the oviduct of 0.5 day pseudopregnant NMRI females.

5

10

20

35

C. Analysis of transgenic mice: Mice were analyzed for the presence of the pCAGGS-C31CNLS-pA transgene by PCR using tail DNA and the primers C31-screen 1 (SEQ ID NO:100) and C31-screen 2 (SEQ ID NO:101) amplifying a fragment of 500 bp. The PCR reaction contained 5 μl PCR buffer (Invitrogen), 2 μl 50 mM MgCl $_2$, 1.5 μl 10 mM dNTP-mix, 2 μl (10 pmol) of each primer, 0.5 μl Taq-polymerase (5 U/ μl) and water to a volume of 50 μl . The program used for the PCR reactions was: 94 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min in 30 cycles.

<u>D. Analysis of C31-Int(CNLS) activity:</u> Founder mice transgenic for the pCAGGS-C31CNLS-pA transgene were crossed to heterozygous C31 reporter mice carrying the C31 reporter construct in the ROSA26 locus (SEQ ID NO:106) (Fig. 8).

Offspring of the crosses were genotyped for the presence of the pCAGGS-C31CNLS-pA transgene by the PCR assay described in section C as well as for the ROSA26-C31 reporter allele by a LacZ-specific PCR assay. The PCR was performed using tail DNA and the primers β-Gal 3 (SEQ ID NO:102) and β-Gal 4 (SEQ ID NO:103) amplifying a fragment of 315 bp. The PCR reaction contained 5 μl PCR buffer (Invitrogen), 2.5 μl 50 mM MgCl₂, 2 μl 10 mM dNTP-mix, 1 μl (10 pmol) of each primer, 0.4 μl Taq-polymerase (5 U/ μl) and water to a volume of 50 μl. The program used for the PCR reactions was: 94 °C for 1 min, 60 °C for 1 min and 72 °C for 1 min in 30 cycles.

Testes from mice carrying the pCAGGS-C31CNLS-pA transgene as well as the reporter locus and from a control mouse carrying the reporter allele only were

dissected. The tissues were imbedded in OCT Tissue freezing medium (Leica/Jung) and frozen in liquid nitrogen. Cryosections were generated from the embedded tissues using a Leica CM3050 cryomicrotome, dried on polylysine-coated slides for 1-4 hours and then stained as follows: Sections were fixed in 0.2 % glutaraldehyde, 5 mM EGTA, 2 mM MgCl₂ in 0.1 M PB (K₂HPO₄/ KH₂PO₄, pH 7.3) for 5 min at room temperature and washed in wash buffer (2 mM MgCl₂, 0.02 % Nonidet-40 in PB in 0.1 M PB) 3 times for 15 min. Then sections were stained in X-Gal-solution (0.6 mg/ ml X-Gal in DMSO, 5 mM potassium hexacyanoferrat III, 5 mM potassium hexacyanoferrat II in LacZ wash buffer) overnigth at 37 °C. After staining sections were washed in 1x PBS twice for 5 min. Dehydration was performed by washing the sections first with 70 %, 96 % and 100 % ethanol for 2 min each, then with a 1:1 mix of ethanol and xylol for 5 min and in the end only with xylol for 5 min. Before taking pictures sections were mounted in Entellan.

15

20

25

30

35

10

E. Results: To identify transgenic founder mice carrying the pCAGGS-C31CNLSpA transgene, 29 mice born from the injection experiment were analyzed for the presence of the transgene. 5 founder mice (3 females and 2 males) were identified. To analyze the activity of the C31-Int(CNLS) recombinase in transgenic mice, 2 of the female founder mice were crossed to heterozygous C31 reporter mice carrying a C31 reporter construct in the ROSA26 locus (Fig. 8). From each of these crosses, one offspring carrying the pCAGGS-C31CNLS-pA transgene as well as the C31 reporter allele was sacrificed. In oder to determine whether pCAGGS-C31CNLS-pA transgenic mice are able to delete an attB/P flanked DNA sequence in the mouse germline, tissue sections from the testes of the sacrificed animals were prepared and stained for ß-galactosidase activity with X-Gal. Fig. 9 shows the result of the staining experiment for one of these mice (A) as well as a control mouse carrying only the reporter allele, but lacking the pCAGGS-C31CNLS-pA transgene (B). Clear staining can be detected in the maturing sperm cells in about 50% of the tubules with the proportion of ßgalactosidase expressing cells ranging between 10 and 100. No staining could be detected for the control mouse. This clearly demonstrates that C31-int-mediated recombination has taken place during spermatogenesis in the pCAGGS-C31CNLSpA transgenic mice. These results show that the C31-int is functional in vivo, in a

WO 02/38613

transgenic mouse system and therefore provides a new tool to introduce specific deletions, inversions or integrations into the mouse germline.

Claims

- 1. A fusion protein comprising
- (a) a recombinase domain comprising a recombinase protein or or a mutant thereof having a recombinase activity similar to that of the corresponding wild-type recombinase and
 - (b) a signal peptide domain linked to said recombinase domain which directs nuclear import of said fusion protein in eucaryotic cells.
- 2. The fusion protein of claim 1, wherein the activity of the fusion protein in eucaryotic cells is significantly higher as compared to that of the wild-type recombinase corresponding to the recombinase of the recombinase domain.
- The fusion protein of claim 1 or 2, wherein the recombinase domain comprises a recombinase protein belonging to the family of large serine recombinases or a mutant thereof, preferably the recombinase domain comprises a recombinase protein selected from the group consisting of bacteriophage ΦC31 integrase (C31-Int), coliphage P4 recombinase, Listeria phage recombinase, bacteriophage R4 Sre recombinase, CisA recombinase, XisF recombinase, transposon Tn4451
 TnpX recombinase and lactococcal bacteriophage TP901-1 recombinase, or a mutant thereof; most preferably the recombinase protein is a C31-Int protein or a mutant thereof.
- 4. The fusion protein of claim 3, wherein the recombinase protein comprises a C31-Int having the amino acid sequence shown in SEQ ID NO:21 or a C-terminal truncated form thereof, said truncated form of the C31-Int preferably comprising amino acid residues of 306 to 613 of SEQ ID NO:21.
- 5. The fusion protein according to any one of claims 1 to 4, wherein the signal peptide domain is derived from yeast GAL4, SKI3, L29 or histone H2B proteins, polyoma virus large T protein, VP1 or VP2 capsid protein, SV40 VP1 or VP2 capsid protein, adenovirus E1a or DBP protein, influenza virus NS1 protein, hepatitis virus core antigen or the mammalian lamin, c-myc, max, c-myb, p53, c-erbA, jun, Tax, steroid receptor or Mx proteins, SV40 T-antigen or other proteins with

known nuclear localisation, preferably the signal peptide domain comprises a peptide which is derived from the SV40 T-antigen.

- 6. The fusion protein according to any one of claims 1 to 5, wherein the signal peptide domain
- (i) has a length of 5 to 74, preferably 7 to 15 amino acid residues, and/or
- (ii) comprises a segment of 6 amino acid residues having at least 2 positively charged basic amino acid residues, said basic amino acid residues being preferably selected from lysine, arginine and histidine.

10

7. The fusion protein of claim 5 or 6, wherein the signal peptide domain comprises a peptide selected from a sequence shown in SEQ ID NOs:24 to 53, preferably the signal peptide comprises the amino acid sequence Pro-Lys-Lys-Lys-Arg-Lys-Val (SEQ ID NO:53).

15

- 8. The fusion protein according to any one of claims 1 to 6, wherein
- (i) the signal peptide domain is linked to the N-terminal or C-terminal of the recombinase domain or is integrated into the recombinase domain, preferably the signal peptide domain is linked to the C-terminal of the recombinase domain; and/or
- (ii) the signal peptide domain is linked to the recombinase domain directly or through a linker peptide, said linker preferably having 1 to 30 essentially neutral amino acid residues.
- The fusion protein of claim 1 comprising the amino acid sequence shown in SEQ ID NO:23.
 - 10. A DNA coding for the fusion protein according to any one of claims 1 to 9.
- 30 11. A vector containing the DNA as defined in claim 10.
 - 12. A microorganism containing the DNA of claim 10 and/or the vector of claim 11.

13. A process for preparing the fusion protein as defined in any one of claims 1 to 9 which comprises culturing a microorganism as defined in claim 11 under conditions suitable for expression of said fusion protein and recovering said fusion protein.

5

- 14. Use of the fusion protein as defined in any one of claims 1 to 9 to recombine DNA molecules, which contain recombinase recognition sequences for the recombinase protein of the recombinase domain, in eucaryotic cells.
- 15. A cell, preferably a mammalian cell containing the DNA sequence of claim 10 in its genome.
 - 16. The cell of claim 15, also containing recognition sequences for the recombinase protein of the recombinase domain in its genome.

- 17. Use of the cell of claim 15 or 16 for studying the function of genes and for the creation of transgenic organisms.
- 18. A transgenic organism, preferably a transgenic non-human mammal containing the DNA sequence of claim 10 in its genome.
 - 19. The transgenic organism of claim 18 also containing recognition sequences for the recombinase protein of the recombinase domain in its genome.
- 25 20. Use of the transgenic organism of claim 18 or 19 for studying gene function at various developmental stages.
- 21. A method for recombining DNA molecules of cells or organisms containing recombinase recognition sequences for the recombinase protein of the recombinase domain as defined in claims 1 to 9, which method comprises supplying the cells or organisms with a fusion protein as defined in claims 1 to 9 or with a DNA sequence of claim 10 and/or a vector of claim 11 which are capable of expressing said fusion protein in the cell or organism.
- 35 22. A method for recombining a DNA molecule containing recognition sequences

for a recombinase protein in a eucaryotic cell, said method comprising contacting the cell with a fusion protein according to claim 1 that recognizes said recognition sequences, wherein the fusion protein catalyzes recombination of the DNA molecule.

- 23. The fusion protein according to any one of claims 1 to 9 which catalyzes recombination at recognition sequences for the recombinase protein.
- 24 A transgenic organism, preferably a transgenic non-human mammal,
 comprising a cell containing a DNA sequence coding for a recombinase fusion protein as defined in claims 1 to 9 and 23 in its genome.

SEQUENCE LISTING

```
<110> Artemis Pharmaceuticals GmbH
  5
      <120> Modified Recombinase
      <130> 012787wo/JH/ml
 10
      <140>
      <141>
      <160> 108
 15
      <170> PatentIn Ver. 2.1
      <210> 1
      <211> 86
      <212> DNA
 20 <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: primer C31-4
 25
      <400> 1
      cgtgacggtc tcgaagccgc ggtgcgggtg ccagggcgtg cccttgggct ccccgggcgc 60
      gtactccacc tcacccatct ggtcca
 30
      <210> 2
      <211> 86
      <212> DNA
      <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: primer C31-5
     <400> 2
     cgtggaccag atgggtgagg tggagtacgc gcccggggag cccaagggca cgccctggcc 60
40
     cacgcaccgc ggcttcgaga ccgtca
     <210> 3
     <211> 90
45
     <212> DNA
     <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: primer C31-6
50
     <400> 3
     gatccagaag cggttttcgg gagtagtgcc ccaactgggg taacctttga gttctctcag 60
     ttgggggcgt agggtcgccg acatgacacg
55
     <210> 4
     <211> 90
     <212> DNA
     <213> Artificial Sequence
60
     <220>
     <223> Description of Artificial Sequence: primer C31-7-2
     <400> 4
65
     gatecgtgte atgteggega ecetaegeee ecaaetgaga gaaetcaaag gttaeeecag 60
     ttggggcact actcccgaaa accgcttctg
```

```
<210> 5
        <211> 7438
        <212> DNA
        <213> Artificial Sequence
        <220>
       <223> Description of Artificial Sequence: vector pRK64
 10
       <400> 5
       cgtcatcacc gaaacgcgcg aggcagctgt ggaatgtgtg tcagttaggg tgtggaaagt 60 ccccaggctc cccagcaggc agaagtatgc aaagcatgca tctcaattag tcagcaacca 120
       ggctccccag caggcagaag tgtgcaaagc atgcatctca attagtcagc aaccatagtc 180
       cegecectaa eteegeceat ecegececta acteegecea gtteegecea tteteegece 240
       catggctgac taatíttttt tatítatgca gaggccgagg ccgcctcggc ctaggaacag 300
       tegacgacac tgcagagacc tacttcacta acaaceggta cagttegtgg accagatggg 360
       tgaggtggag tacgcgccg gggagcccaa gggcacgcc tggcacccgc accgcggctt 420
       cgagaccgtc acgaataact tcgtatagca tacattatac gaagttataa gcttgcatgc 480
       ctgcaggtcg gccgccacga ccggccggcc ggtgccgcca ccatcccctg acccacgccc 540 ctgacccctc acaaggagac gaccttccat gaccgagtac aagcccacgg tgcgcctcgc 600
 20
       caccegegae gaegteece gggeegtaeg caccetegee geegegtteg cegaetaeee 660
       cgccacgcgc cacaccgtcg acccggaccg ccacatcgag cgggtcaccg agctgcaaga 720
       actottocto acgogogtog ggotogacat oggoaaggtg tgggtogogg acgaoggogo 780
      cgcggtggcg gtctggacca cgccggagag cgtcgaagcg ggggcggtgt tcgccgagat 840 cggcccgcgc atggccgagt tgagcggttc ccggctggcc gcgcagcaac agatggaagg 900
 25
       cetectggeg cegeacegge ceaaggagee egegtggtte etggeeaeeg teggegtete 960
      gcccgaccac cagggcaagg gtctgggcag cgccgtcgtg ctccccggag tggaggcggc 1020
      cgagcgcgcc ggggtgcccg cetteetgga gacctccgcg ccccgcaacc tccccttcta 1080
      cgageggete ggetteaceg teacegeega egtegagtge eegaaggace gegegacetg 1140
 30
      gigcatgace egcaageeeg gigceigaeg eccgeeceae gaccegeage geeegaeega 1200
      aaggagegea egaceceatg geteegaceg aageegacee gggeggeeee geegaceeeg 1260 caccegeeee egaggeeeae egactetaga ggateataat eageeataee acatttgtag 1320
      aggttttact tgctttaaaa aacctcccac acctccccct gaacctgaaa cataaaatga 1380
      atgcaattgt tgttgttaac ttgtttattg cagcttataa tggttacaaa taaagcaata 1440
 35
      gcatcacaaa tttcacaaat aaagcatttt tttcactgca ttctagttgt ggtttgtcca 1500
      aactcatcaa tgtatcttat catgtctgga tccgtgtcat gtcggcgacc ctacgccccc 1560
      aactgagaga actcaaaggt taccccagtt ggggcactac teccgaaaac cgcttetgga 1620
      tocataactt cgtatagcat acattatacg aagttatacc gggccaccat ggtcgcgagt 1680
      agettggeae tggeegtegt tttacaaegt egtgaetggg aaaaceetgg egttaceeaa 1740
 40
      cttaategee ttgeageaca tecceettte geeagetgge gtaatagega agaggeeege 1800
      accgategee etteceaaca gttgegeage etgaatggeg aatggegett tgeetggttt 1860
      ccggcaccag aagcggtgcc ggaaagctgg ctggagtgcg atcttcctga ggccgatact 1920
      gtcgtcgtcc cctcaaactg gcagatgcac ggttacgatg cgcccatcta caccaacgta 1980 acctatccca ttacggtcaa tccgccgttt gttcccacgg agaatccgac gggttgttac 2040
45
      tegeteacat ttaatgttga tgaaagetgg ctacaggaag gecagaegeg aattatttt 2100
      gatggcgtta actcggcgtt tcatctgtgg tgcaacgggc gctgggtcgg ttacggccag 2160 gacagtcgtt tgccgtctga atttgacctg agcgcatttt tacgcgccgg agaaaaccgc 2220
      ctcgcggtga tggtgctgcg ttggagtgac ggcagttatc tggaagatca ggatatgtgg 2280
      cggatgageg geatttteeg tgaegteteg ttgetgeata aaccgaetae acaaatcage 2340
50
      gatttccatg ttgccactcg ctttaatgat gatttcagcc gcgctgtact ggaggctgaa 2400
      gttcagatgt gcggcgagtt gcgtgactac ctacgggtaa cagtttcttt atggcagggt 2460
      gaaacgcagg tcgccagcgg caccgcgcct ttcggcggtg aaattatcga tgagcgtggt 2520
      ggttatgccg atcgcgtcac actacgtctg aacgtcgaaa acccgaaact gtggagcgcc 2580
      gaaatcccga atctctatcg tgcggtggtt gaactgcaca ccgccgacgg cacgctgatt 2640
55
      gaagcagaag cetgegatgt eggttteege gaggtgegga ttgaaaatgg tetgetgetg 2700
      ctgaacggca agccgttgct gattcgaggc gttaaccgtc acgagcatca tcctctgcat 2760
      ggtcaggtca tggatgagca gacgatggtg caggatatcc tgctgatgaa gcagaacaac 2820
      titaacgccg tgcgctgttc gcattatccg aaccatccgc tgtggtacac gctgtgcgac 2880
     cgctacggcc tgtatgtggt ggatgaagcc aatattgaaa cccacggcat ggtgccaatg 2940
60
     aatcgtctga ccgatgatcc gcgctggcta ccggcgatga gcgaacgcgt aacgcgaatg 3000
     gtgcagcgcg atcgtaatca cccgagtgtg atcatctggt cgctggggaa tgaatcaggc 3060
     cacggegeta ateaegaege getgtatege tggateaaat etgtegatee tteeegeeeg 3120
     gtgcagtatg aaggcggcgg agccgacacc acggccaccg atattatttg cccgatgtac 3180
     gegegegtgg atgaagacea geeetteeeg getgtgeega aatggteeat caaaaaatgg 3240
65
     ctttcgctac ctggagagac gcgcccgctg atcctttgcg aatacgccca cgcgatgggt 3300
     aacagtettg geggtttege taaatactgg caggegttte gteagtatee eegtttacag 3360
     ggcggcttcg tctgggactg ggtggatcag tcgctgatta aatatgatga aaacggcaac 3420
```

| | | | | 3 | | | |
|----|-------------|---|--------------------|------------------------------|-----------------------|--------------------------|-------|
| | ccgtggtcgg | cttacggcgg | r tgattttggd | gatacqccqa | acgategeca | gttctgtatg | 3480 |
| | ~~vggcccqq | LCLCLUCCUA | CCCCacacac | T Catecoace | + ~~~~~~~~~ | | 25.42 |
| | | - LUCAULUCCO | LIFATCCOOC | 1 0222002400 | | | |
| _ | | gegataatqa | uctoctocar | : rogatogtog | - cactaastaa | + | 2000 |
| 5 | 3~~~g~gg~g | aagtqcctt | - qualulcuci | . ccacaaddra | aacaattaat | +~~~~+ | 2722 |
| | gaactactgc | ayccuyayay | caccaaacaa | l CtCtddctca | cartacecet | | 2200 |
| | aucycyaccy | Caryyrcaga | auccuaacac | : afcadedect | addeaddeadta | ~~~+ ~+ ~~~ | 2040 |
| | gaaaacctca | gtgtgacgct | ccccaccaca | r teceaegeee | tocogcatet | gaccaccage | 3840 |
| | gaaatggatt | tttgcatcga | gctgggtaat | aagcattaac | aatttaaccc | ccagtcagge | 3900 |
| 10 | tttctttcac | agatgtggat | toocoataaa | aaacaactoo | taacccaaccy | gcgcgatcag | 3960 |
| | ttcacccgtg | caccactaga | taacgacatt | . aaacaaccyc | 230003CCCC | cattgaccct | 4020 |
| | aacgcctggg | togaacgotg | gaagggggg | ngccattacc | aagcgacccg | agcgttgttg | 4080 |
| | cagtgcacgg | cagatacact | tactaataca | r gyccartact r gtgctgette | aggeegaage | cgcgtggcag | 4140 |
| | catcagggga | aaaccttatt | tatcagegg | , gegetgatta | cgaccgccca | tagtggtcaa | 4200 |
| 15 | atggcgatta | ccattaatat | tgaagtgggg | addacceacc | ggartgargg | gcggattggc | 4260 |
| | ctgaactgcc | agctggcgca | aataacaaa | cacataaac | cgcatccggc | agggccgcaa | 4320 |
| | gaaaactatc | ccaaccacct | tactaccacc | tattttaace | ggeteggatt | gccattgtca | 4380 |
| | gacatgtata | ccccatacat | cttcccaac | r deservant | tacastacas | gacgcgcgaa | 4440 |
| | ttgaattatg | acceseace | ataacacaaa | gadaacygtt | tgegergegg | ccgctacagt | 4500 |
| 20 | caacagcaac | tgatggaaac | Canccatege | gacetecage | ccaacatcag | aggcacatgg | 4560 |
| | ctgaatatcg | accotttcca | tatagagatt | gatagagaga | acgcggaaga | aggcacatgg cccgtcagta | 4620 |
| | toggoggaat | tccagctgag | caccaatcac | taccattaca | actectggag | gtgtcaaaaa | 4680 |
| | taataataac | cadacadada | ggatctttgt | dascarract | tacttetete | gtgtcaaaaa gtgtgacata | 4740 |
| | attqqacaaa | ctacctacag | agatttaaag | ctctaaggt | cacccccgcg | ttttaagtgt | 4800 |
| 25 | ataatgtgtt | aaactactga | ttctaattgt | ttatatatat | tagattaga | cctatggaac | 4860 |
| | tgatgaatgg | gagcagtogt | ggaatgccag | atcoacacat | cayatteccaa | attgatgagt | 4920 |
| | ttggacaaac | cacaactaga | atgcagtgaa | accoayacat | tatttataa | atttgtgatg | 4980 |
| | ctattqcttt | atttgtaacc | attataaget | gcaataaag | actigigaa | acceattgca | 5040 |
| | ttcattttat | gtttcaggtt | Cadddddaad | tataaaaaa | tttttaacaac | aacaattgca aagtaaaacc | 5100 |
| 30 | tctacaaatq | tagtataget | gattatgate | tacaaccac | ggggtggtg | atacgcctat | 2160 |
| | ttttataggt | taatotcato | ataataatoo | tttcttagag | gggccccgcg | acttttcggg | 5220 |
| | gaaatgtgcg | cggaacccct | atttatttat | ttttctaaat | greaggregge | atgtatccgc | 5280 |
| | tcatgagaca | ataaccctga | taaatgette | aataatatta | acattcaaat | agtatgagta | 5340 |
| | ttcaacattt | ccatateace | cttattccct | tttttaaaa | addadygady | cctgtttttg | 5400 |
| 35 | ctcacccaga | aacoctooto | aaaataaaa | atactasaas | tooattaaat | gcacgagtgg | 5460 |
| | gttacatcga | actogatete | aacagcggta | acgccgaaga | ccayttyggt | cccgaagtgg | 5520 |
| | gttttccaat | gatgaggact | tttaaagttc | tactatataa | gageetatta | tcccgtattg | 5580 |
| | acgccgggca | agaggaacte | gatcaccaca | tacactatto | toacastas | ttggttgagt | 5640 |
| | actcaccagt | cacagaaaag | catcttacco | ataggetaec | actananan | ttatgcagtg | 5700 |
| 40 | ctgccataac | catgagtgat | aacactgcgg | ccaacttact | totangagaa | atcggaggac | 5/60 |
| | cgaaggagct | aaccoctttt | ttgcacaaca | taggagatas | totgacaacg | cttgatcgtt | 5820 |
| | gggaaccqqa | gctgaatgaa | occataccaa | egggggacca | tgraactege | atgcctgtag | 5880 |
| | caatggcaac | aacgttgcgc | aaactattaa | ctancasact | acttacteta | gcttcccggc | 5940 |
| | -ucaattaat | auactodato | uauucuuara | aanttaaaa | ~~~~ ~++ ~ | | |
| 45 | ttccggctgg | ctqqtttatt | gctgataaat | ctggaggggg | tgagggtggg | tctcgcggta | 6100 |
| | | actududcca | ua luulaane | CCTCCCGTat | COTTORTEDEO | + | C100 |
| | ggagccaggc | aactatqqat | gaacgaaata | gacagatege | tagastsaat | goot on other | COAO |
| | cuagcatty | ytaactytca | gaccaagttt | actcatatat | actttagatt | antttanna. | 6300 |
| | cccaccccca | atttaaaagg | atctaggtga | agatectitt | trataatoto | atracraaaa | 6260 |
| 50 | - cccccaacg | Lyayttttcg | ttccactgag | Cotcagaccc | cotagaaaa | atrasarrat | 6420 |
| | | CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | ctococotaa | tetactactt | TC222C222 | 222002000 | CADO |
| | caccagogge | ggtttgtttg | ccggatcaag | agetaceaac | tetttteee | 220012201- | CEAO |
| | getteageag | aqcqcaqata | ccaaatacto | teettetaat | at agoogt ag | ++-~~~ | CCOO |
| | accicaayaa | Cicigtagea | ccccctacat | acctedetet | actastocta | ttaccacter. | CCCO |
| 55 | orgergeeag | Lygogataag | tcatatetta | CCGGGGttgGG | rtraanarna | taattaaaaa | 6720 |
| | acaaggcgca | gcggccgggc | tgaacggggg | atteatacae | acadeceade | ttaaaaaaa | 6700 |
| | egacctacac | cgaactgaga | tacctacage | Qtqaqctatq | adaaadcdcc | acactteces | COAN |
| | aagggagaaa | ggcggacagg | tatcccctaa | acaacsaaat | cadaagegee | gaggggagg | 6000 |
| | gggagettee | aqqqqqaaac | acctagtate | tttatactcc | tataaaattt | accessortet | COCO |
| 60 | gacttgagcg | tcgatttttg | tgatgctcgt | Cadadadaca | gacctataa | aaaaaccccc | 7020 |
| | geaucycygc | cttttatgg | LLCCLqqcct | rrractaacc | ttttactcac | atattatta | 7000 |
| | ctgcgttatc | ccctgattct | gtggataacc | gtattaccgc | ctttgactaa | acguette. | 7140 |
| | ctcgccgcag | ccgaacgacc | gagegeageg | agtcantgag | casaassaca | geryaraceg | 7200 |
| | caatacgcaa | accocctctc | cccqcacatt | aaccasttc= | ttaatoosoo | tagaagagagaa | 7260 |
| 65 | ggtttcccga | ctggaaagcg | ggcagtgagc | acaacacaat | taatotosot | tageterate | 7220 |
| | accayycacc | ccaggettta | cactttatac | ttacaaataa | tatottotot | acaattataa | 7300 |
| | gcggataaca | atttcacaca | ggaaacagct | atgaccatga | ttacccceec | gyaarrytyd ctaacaaa | |
| | | | | | gooday | - eggegeg | 7438 |

```
<210> 6
      <211> 45
      <212> DNA
      <213> Artificial Sequence
     <223> Description of Artificial Sequence: primer C31-1
 10
      <400> 6
      ataagaatgc ggccgcccga tatgacacaa ggggttgtga ccggg
                                                                        45
 15
      <210> 7
      <211> 40
      <212> DNA
     <213> Artificial Sequence
20 <220>
     <223> Description of Artificial Sequence: primer C31-3
     ataagaatgc ggccgcatcc gccgctacgt cttccgtgcc
                                                                        40
25
     <210> 8
     <211> 24
     <212> DNA
30
     <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: primer C31-8
35
     <400> 8
     cccgttggca ggaagcactt ccgg
                                                                        24
     <210> 9
40
     <211> 55
     <212> DNA
     <213> Artificial Sequence
     <220>
45
     <223> Description of Artificial Sequence: primer C31-9
     ggatcctcga gccgcgggcg gccgcctacg ccgctacgtc ttccgtgccg tcctg
50
     <210> 10
     <211> 5711
     <212> DNA
     <213> Artificial Sequence
55
     <220>
     <223> Description of Artificial Sequence: vector
           pCMV-C31-Int(wt)
60
     <400> 10
    aaacagtccg atgtacgggc cagatatacg cgttgacatt gattattgac tagttattaa 60
    tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg cgttacataa 120
     cttacggtaa atggeeegee tggetgaeeg eecaacgaee eeegeeeatt gaegteaata 180
    atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca atgggtggac 240
65 tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc aagtacgccc 300
    cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta catgacctta 360
    tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac catggtgatg 420
```

| | cggttttggc | agtacatcaa | tgggcgtgga | tagoggtttg | actcacqqqq | atttccaagt | 480 |
|----|------------|-------------|-------------|-------------|--------------------------|------------|------|
| | ctccacccca | ttgacgtcaa | tgggagtttg | ttttggcacc | aaaatcaaco | ggactttcca | 540 |
| | aaatgtcgta | acaactccgc | cccattgacg | caaatgggcg | ataggcatat | acqqtqqqaq | 600 |
| _ | gtctatataa | gcagagctct | ctggctaact | agagaaccca | ctocttacto | gettategaa | 660 |
| 5 | attaatacga | ctcactatag | ggagacccaa | gctgactcta | gacttaatta | agcattagga | 720 |
| | tgagtactcc | ctctcaaaag | cgggcatgac | ttctqcqcta | agattotcag | tttccaaaaa | 780 |
| | cgaggaggat | ttgatattca | cctggcccgc | ggtgatgcct | ttgagggtgg | ccacatccat | 840 |
| | ctggtcagaa | aagacaatct | ttttattatc | aagcttgagg | tgtggcaggc | ttgagatctg | 900 |
| | gccatacact | tgagtgacat | tgacatccac | tttgcctttc | tctccacagg | totccactcc | 960 |
| 10 | cagggcggcc | gcccgatatg | acacaagggg | ttataaccaa | ggtggacacg | tacccccccc | 1020 |
| | cttacgaccg | tcagtcgcgc | gagegegaga | attegagege | agcaagccca | acascade | 1020 |
| | gtagcgccaa | cqaaqacaaq | gcggccgacc | ttcagcgcga | agtcgagcgc | ascadadacc | 1140 |
| | ggttcaggtt | categggcat | ttcagcgaag | caccaaacsc | gtcggcgttc | daascaacaa | 1200 |
| | agcqcccqqa | gttcgaacgc | atcctgaacg | aatoccococ | cgggcggctc | aacatoatca | 1260 |
| 15 | ttgtctatga | catateacae | ttctcacacc | tgaaggtcat | ggacgcgatt | ccaattatct | 1320 |
| | cggaattgct | caccetaaac | gtgacgattg | tttccactca | ggaaggcgtc | ttccaccaa | 1380 |
| | gaaacgtcat | ggacctgatt | cacctgatta | tacaactcaa | cgcgtcgcac | aaadaatctt | 1440 |
| | cgctgaagtc | ggcgaagatt | ctcgacacga | agaacettca | gcgcgaattg | aaagaateet | 1500 |
| | toggogggaa | ggcgccttac | ggcttcgagc | ttatttcaga | gacgaaggag | atcacacaca | 1560 |
| 20 | acggccgaat | ggtcaatgtc | gtcatcaaca | agettgegea | ctcgaccact | cccttacca | 1620 |
| | gaccettega | gttcgagece | gacgtaatcc | aataataata | gcgtgagatc | aagacgcaca | 1680 |
| | aacaccttcc | cttcaagccg | ggcagtcaag | ccccattca | cccgggcagc | atcacaaaaa | 1740 |
| | tttgtaagcg | catggacgct | gacgccgtgc | caacccaaaa | cgagacgatt | gggaagaaga | 1800 |
| | ccgcttcaag | cgcctgggac | ccggcaaccg | ttatocoaat | ccttcgggac | ccacatatta | 1860 |
| 25 | cgggcttcgc | cgctgaggtg | atctacaaga | agaagccgga | cggcacgccg | accacgaaga | 1920 |
| | ttgagggtta | ccgcattcag | cgcgacccga | tcacactcca | gccggtcgag | cttgattgcg | 1980 |
| | gaccgatcat | cgagcccgct | gagtggtatg | agetteagge | gtggttggac | aacaaaaaaa | 2040 |
| | gcggcaaggg | gctttcccgg | gggcaagcca | ttctgtccgc | catggacaag | ctgtactgcg | 2100 |
| | agtgtggcgc | cgtcatgact | tcgaagcgcg | gggaagaatc | gatcaaggac | tcttaccact | 2160 |
| 30 | gccgtcgccg | gaaggtggtc | gacccgtccg | cacctgggca | gcacgaaggc | acqtqcaacq | 2220 |
| | tcagcatggc | ggcactcgac | aagttcgttg | cggaacgcat | cttcaacaag | atcaggcacg | 2280 |
| | ccgaaggcga | cgaagagacg | ttggcgcttc | tgtgggaagc | cqcccqacqc | ttcggcaage | 2340 |
| | tcactgaggc | gcctgagaag | agcggcgaac | gggcgaacct | tgttgcggag | cqcqccqacq | 2400 |
| 25 | ccctgaacgc | ccttgaagag | ctgtacgaag | accgcgcggc | aggcgcgtac | gacggacccg | 2460 |
| 35 | ttggcaggaa | gcacttccgg | aagcaacagg | cagcgctgac | gctccggcag | caaggggggg | 2520 |
| | aagagcggct | tgccgaactt | gaagecgccg | aagccccgaa | gcttcccctt | gaccaatggt | 2580 |
| | tccccgaaga | cgccgacgct | gacccgaccg | gccctaagtc | gtggtgggg | cacacatcaa | 2640 |
| | tagacgacaa | gcgcgtgttc | gtcgggctct | tcgtagacaa | gatcqttqtc | acqaaqtcqa | 2700 |
| 40 | ctacgggcag | ggggcaggga | acgcccatcg | agaagcqcqc | ttcgatcacg | tagacaaac | 2760 |
| 40 | cgccgaccga | cgacgacgaa | gacgacgccc | aggacggcac | ggaagacgta | gcggcgtagg | 2820 |
| | cggcgcccgg | gctcgagatc | caggcgcgga | tcaataaaag | atcattatt | tcaatagatc | 2880 |
| | tgtgtgttgg | ttttttgtgt | gccttggggg | agggggaggc | cagaatgagg | cgcggccaag | 2940 |
| | ggggaggggg | aggccagaat | gaccttgggg | gagggggagg | ccagaatgac | cttgggggag | 3000 |
| 45 | ggggaggcca | gaatgaggcg | cdcccccddd | taccgagete | gaattcactg | gccgtcgttt | 3060 |
| 73 | cacaacgtcg | tgactgggaa | aaccetggeg | ttacccaact | taatcgcctt | gcagcacatc | 3120 |
| | tacacacac | cagetggegt | aatagcgaag | aggcccgcac | cgatcgccct | tcccaacagt | 3180 |
| | cycycaycc: | gaacggegaa | tggcgcctga | tgcggtattt | tctccttacg | catctgtgcg | 3240 |
| | gcatttcaca | cegeatatgg | tgcaetetea | gtacaatctg | ctctgatgcc | gcatagttaa | 3300 |
| 50 | Categorita | acacccgcca | acacecycig | acgegeeetg | acgggcttgt | ctgctcccgg | 3360 |
| | Catcatcacc | cayacaaycc | grgacegree | ccgggagctg | catgtgtcag | aggttttcac | 3420 |
| | atotoatoat | antantagtt | agacgaaagg | gcctcgtgat | acgectattt | ttataggtta | 3480 |
| | daaccoctat | ttatttattt | ttataaata | caggraggeac | ttttcgggga | aatgtgcgcg | 3540 |
| | aaccctcata | aatoottoaa | taatattaa | actcaaatat | gtatccgctc tatgagtatt | argagacaar | 3600 |
| 55 | atateacect | tattcccttt | tttaararaat | tttaattaa | tgtttttgct | caacatttcc | 3660 |
| | Cactagtass | anteseanat | actanacata | anthemates | acgagtgggt | cacccagaaa | 3720 |
| | togateteaa | caccactaaa | atcetteaca | agitigggige | cgaagaacgt | tacatcgaac | 3/80 |
| | tgaggagttt | taaaatteta | stateteese | greetettete | ccgtattgac | tttccaatga | 3840 |
| | agcaactcgc | tcaccacata | cactattoto | eggiatiate | ggttgagtac | yeegggcaag | 3900 |
| 60 | Cadaaaadca | tcttaccat | caccatteres | taaaagactt | atgcagtgct | tcaccagtca | 3960 |
| | taaataataa | cactacaggat | aacttactta | tanananant | cggaggaccg | yccataacca | 4020 |
| | Coacttttt | acacaacata | aggetageta | taactacat | tgatcgttgg | aaggagctaa | 4080 |
| | tgaatgaage | cataccasac | ascasacata | acaccaccat | gcctgtagca | yaaccygagc | 4240 |
| , | Cattacacaa | actattaact | adcasactsc | ttactctacc | ttcccggcaa | aryycaacaa | 4200 |
| 65 | actogatoga | ggcggataaa | gttgcagga | cactteteee | ctcggccctt | cogcotacat | 4200 |
| | ggtttattgc | tgataaatct | ggagccoota | agcatanata | tcgcggtatc | attocaccac | 425U |
| | tggggccaga | togtaagccc | tecegtates | tagttatcta | cacgacgggg | antragran | 4700 |
| | | | 3 | | | ageouggeda | 1110 |
| | | | | | | | |

PCT/EP01/12975

WO 02/38613 ctatggatga acgaaataga cagatcgctg agataggtgc ctcactgatt aagcattggt 4500 aactgtcaga ccaagtttac tcatatatac tttagattga tttaaaactt catttttaat 4560 ttaaaaggat ctaggtgaag atcettttg ataateteat gaccaaaate cettaacgtg 4620 agttttcgtt ccactgagcg tcagaccccg tagaaaagat caaaggatct tcttgagatc 4680 cttttttct gcgcgtaatc tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg 4740 tttgtttgcc ggatcaagag ctaccaactc tttttccgaa ggtaactggc ttcagcagag 4800 cgcagatacc aaatactgtc cttctagtgt agccgtagtt aggccaccac ttcaagaact 4860 ctgtagcacc gcctacatac ctcgctctgc taatcctgtt accagtggct gctgccagtg 4920 gcgataagtc gtgtcttacc gggttggact caagacgata gttaccggat aaggcgcagc 4980

10 ggtcgggctg aacggggggt tcgtgcacac agcccagctt ggagcgaacg acctacaccg 5040 aactgagata cctacagcgt gagctatgag aaagcgccac gcttcccgaa gggagaaagg 5100 cggacaggta tccggtaagc ggcagggtcg gaacaggaga gcgcacgagg gagcttccag 5160 ggggaaacgc ctggtatctt tatagtcctg tcgggtttcg ccacctctga cttgagcgtc 5220 gatttttgtg atgetegtea ggggggggga geetatggaa aaaegeeage aaegeggeet 5280 ttttaeggtt cetggeettt tgetggeett ttgeteacat gttettteet gegttateee 5340 15 ctgattctgt ggataaccgt attaccgcct ttgagtgagc tgataccgct cgccgcagcc 5400 gaacgaccga gcgcagcgag tcagtgagcg aggaagcgga agagcgccca atacgcaaac 5460 cgcctctccc cgcgcgttgg ccgattcatt aatgcagctg gcacgacagg tttcccgact 5520 ggaaagcggg cagtgagcgc aacgcaatta atgtgagtta gctcactcat taggcacccc 5580 20

aggetttaca etttatgett eeggetegta tgttgtgtgg aattgtgage ggataacaat 5640 ticacacagg aaacagctat gaccatgatt acgccaagct agcccgggct agcttgcatg 5700 cctgcaggtt t

25 <210> 11 <211> 69 <212> DNA <213> Artificial Sequence

30

<223> Description of Artificial Sequence: primer C31-2-2

<400> 11 tagaatteeg etegagagte taaacettee tettettett aggegeeget acgtetteeg 60 35 tgccgtcct

<210> 12 <211> 5723 40 <212> DNA <213> Artificial Sequence

<220> <223> Description of Artificial Sequence: vector 45 pCMV-C31-Int(CNLS)

cctgcaggtt taaacagtcc gatgtacggg ccagatatac gcgttgacat tgattattga 60 ctagttatta atagtaatca attacggggt cattagttca tagcccatat atggagttcc 120 50 gcgttacata acttacggta aatggcccgc ctggctgacc gcccaacgac ccccgcccat 180 tgacgtcaat aatgacgtat gttcccatag taacgccaat agggactttc cattgacgtc 240 aatgggtgga ctatttacgg taaactgccc acttggcagt acatcaagtg tatcatatgc 300 caagtacgcc ccctattgac gtcaatgacg gtaaatggcc cgcctggcat tatgcccagt 360 acatgacctt atgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta 420 ccatggtgat gcggttttgg cagtacatca atgggcgtgg atagcggttt gactcacggg 480 gatttccaag totccaccc attgacgtca atgggagttt gttttggcac caaaatcaac 540 gggactttcc aaaatgtcgt aacaactccg ccccattgac gcaaatgggc ggtaggcgtg 600 tacggtggga ggtctatata agcagagctc tctggctaac tagagaaccc actgcttact 660

ggcttatcga aattaatacg actcactata gggagaccca agctgactct agacttaatt 720 60 aagcgttggg gtgagtactc cctctcaaaa gcgggcatga cttctgcgct aagattgtca 780 gtttccaaaa acgaggagga tttgatattc acctggcccg cggtgatgcc tttgagggtg 840 gccgcgtcca tctggtcaga aaagacaatc tttttgttgt caagcttgag gtgtggcagg 900 cttgagatct ggccatacac ttgagtgaca ttgacatcca ctttgccttt ctctccacag 960

gtgtccactc ccagggcggc cgcccgatat gacacaaggg gttgtgaccg gggtggacac 1020 gtacgcgggt gcttacgacc gtcagtcgcg cgagcgcgag aattcgagcg cagcaagccc 1080 agcgacacag cgtagcgcca acgaagacaa ggcggccgac cttcagcgcg aagtcgagcg 1140 cgacgggggc cggttcaggt tcgtcgggca tttcagcgaa gcgccgggca cgtcggcgtt 1200

| | | | | 7 | | | |
|------------|---|-------------|------------|--------------------------|----------------|------------|---------|
| | cgggacggcg | gagegeeegg | agttcgaacg | catcctgaac | gaatgeegeg | ccaaacaact | 1260 |
| | caucatgatt | accycciaty | acutotcoco | CEECECACAC | ctnaanntoa | taasaaast | 1220 |
| | tccgattqtc | tcggaattgc | tegeetaaa | catalogogo | atttant. | Lygacycyat | 1320 |
| | cttccaacaa | anasacates | tagaaataat | taracturate | grirecacte | aggaaggcgt | 1380 |
| 5 | caaacaatct | ggaaacgtca | cygaccigat | teacetgatt | argeggeteg | acgcgtcgca | 1440 |
| - | aaaaaaaat | tcgctgaagt | cygcgaagat | tctcgacacg | aagaaccttc | agcgcgaatt | 1500 |
| | gggcggcac | yccyycygaa | addcdcctta | caacttcaaa | cttatttcaa | 2020022000 | 1 5 6 6 |
| | gaccacgcgc | aacyyccyaa | tuutcaatot | COTCATCAAC | aagettgege | 201002000 | 1620 |
| | COOCCCCACC | ggacccccccq | aquicqaqcc | coacotaate | caataataat | aacatasast | 1600 |
| 10 | caugucgcac | adacaccccc | CCLTCAAGCC | ggggagtgaa | accaccatta | 2000000000 | 1710 |
| 10 | catcacgggg | ctttgtaagc | gcatggacgc | tgacgccgtg | CCGSCCCGGG | accegggeag | 1000 |
| | tgggaagaag | accgcttcaa | acacctuada | cccaacaaca | attatacasa | gcgagacgat | 1000 |
| | cccccctatt | gcgggcttcg | ccactagga | ceeggeaace catctacaac | greatgegaa | teetteggga | T890 |
| | рассаслаал | attgagggtt | accounter | gatttataag | aagaagccgg | acggcacgcc | 1920 |
| | acttasttas | attgagggtt | accycatica | gegegaeeeg | atcacgctcc | ggccggtcga | 1980 |
| 15 | geergarryc | ggaccgatca | regageeege | tgagtggtat | gagcttcagg | cgtggttgga | 2040 |
| 10 | cggcaggggg | cycggcaagg | adctttccca | adadcaaacc | attetateea | ccatagagaa | 2100 |
| | gergracige | gagiging | ccgtcatgac | ttcgaagcgc | ggggaagaat | coatcaagga | 2160 |
| | cecetaccyc | Lycegregee | qqaaqqtqqt | caacccatcc | acacctagae | BUCSCUSSUU | 2220 |
| | cacgiglaac | greageargg | cggcactcga | caagttcgtt | gcggaacgca | tetteaacaa | 2280 |
| | gaccayycac | gccgaaggcg | acqaaqaqac | attaacactt | ctataaaaa | CCCCCCCCCC | 2210 |
| 20 | cttcggcaag | ctcactgagg | cacctaagaa | dadcadcass | caaacassca | ttattagaca | 2340 |
| | gcgcgccgac | gccctgaacg | cccttgaaga | actatacaaa | and a constant | ctgttgegga | 2400 |
| | cgacggaccc | attagcagga | agcacttccg | googcacgaa | gaccgcgcgg | caggegegta | 2460 |
| | acasaaaaaa | gttggcagga | ttaaaaaaat | yaaycaacag | gcagcgctga | cgctccggca | 2520 |
| | traccastra | gaagagcggc | Ligeegaaet | rgaagccgcc | gaagccccga | agcttcccct | 2580 |
| 25 | agaccaatgg | ttccccgaag | acgccgacgc | tgacccgacc | ggccctaagt | cgtggtgggg | 2640 |
| 23 | gegegegeea | gragacgaca | agcgcgtgtt | catcaaactc | ttegtagaca | adatedttat | 2700 |
| | cacyaayiig | actacgggca | qqqqqcaqqq | aacocccatc | gagaagegeg | cttcgatcac | 2760 |
| | gryggcgaag | ccyccgaceg | acgacgacga | agacgacgcc | caddacddca | cadaacat | 2820 |
| | ageggegeet | aayaagaaga | ggaaggttta | gactctcgag | atecaggege | anatraataa | 2880 |
| | aayattatta | LLLLCaatag | atctqtqtqt | taattttta | tatacettaa | addadadaa | 2010 |
| 30 | ggccagaatg | aggcgcggcc | aagggggagg | addadaccad | aatgaccttg | adadadada | 2000 |
| | aggccagaat | gaccttgggg | asaaaaaaaa | ccaraatrar | aacgaccccg | gggggggg | 3000 |
| | ctcgaattca | ctaaccatca | ++++202200 | taataaataa | gcgcgccccc | gggtaccgag | 3060 |
| | acttaatcoc | ctggccgtcg | atecacatty | Logigaeigg | gaaaaccctg | gcgttaccca | 3120 |
| | caccaatege | cttgcagcac | arceceettt | cgccagctgg | cgtaatagcg | aagaggcccg | 3180 |
| 35 | ttttataatt | ccttcccaac | agrigedead | cctgaatggc | gaatggcgcc | tgatgcggta | 3240 |
| 55 | ctcccccccc | acgcatctgt | gcggtatttc | acaccgcata | tggtgcactc | tcagtacaat | 3300 |
| | cigcicigal | gccgcatagt | taagccagcc | CCGacaccca | CCAACACCCC | ctascacaca | 3360 |
| | cegacygyce | rgrergeree | cggcatccgc | ttacagacaa | actataacca | tetecaaaaa | 3/120 |
| | cigcalgigi | cayaggtttt | caccqtcatc | accgaaacgc | gcgagacgaa | agggeeteat | 7/80 |
| 40 | gatacyccia | LLLLCatagg | ttaatgtcat | gataataatg | gtttcttaga | catcagatag | 3540 |
| 40 | cacttttcgg | ggaaatgtgc | gcggaacccc | tatttottta | tttttctaaa | tacattcasa | 3500 |
| | tatgtatccg | ctcatgagac | aataacccto | ataaatactt | castastatt | cacacccaaa | 3000 |
| | gagtatgagt | attcaacatt | tecatateae | ccttattccc | tttttt | yaaaaaygaa | 3660 |
| | tectatttt | actcacccac | aaaaataat | ~~~~ | cccccgcgg | cattttgcct | 3/20 |
| | tacacaaata | gctcacccag | anacyceyye | yaaagtaaaa | gatgctgaag | atcagttggg | 3780 |
| 45 | CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | ggttacatcg | aactygatet | caacagcggt | aagatccttg | agagttttcg | 3840 |
| | atacastat | cgttttccaa | tgatgagcac | ttttaaagtt | ctgctatgtg | gcgcggtatt | 3900 |
| | attecegeace | gacgccgggc | aagagcaact | cggtcgccgc | atacactatt | ctcagaatga | 3960 |
| | cruggingag | tactcaccag | tcacagaaaa | gcatcttacq | gatggcatga | cantaanana | 4020 |
| | actacycage | gctgccataa | ccatgagtga | taacactgcg | accaacttac | ttctcacaac | 1000 |
| 50 | gateggagga | ccgaaggage | taaccgcttt | tttqcacaac | atgggggate | atotaactoo | 4140 |
| 50 | ccccyatcyt | rgggaaccgg | agctgaatga | agecatacea | aacdacdadc | atascaccac | 4200 |
| | gatgcctgta | gcaatggcaa | caacattaca | caaactatta | actoocoaac | tacttactct | 1260 |
| | agcttcccgg | caacaattaa | tagactggat | agaggggat | accegeegaac | cacccactct | 4200 |
| | gcgctcgacc | cttccggctg | actaatttat | tactactac | totagecycay | gaccaccicc | 4320 |
| | atctcacaat | atcattacaa | googgoodac | ryctyataaa | cciggageeg | grgagcgrgg | 4380 |
| 55 | Ctacaccacc | atcattgcag | cactggggcc | agalggtaag | ccctcccgta | tegtagttat | 4440 |
| •• | tacatasata | gggagtcagg | caactatgga | tgaacgaaat | agacagatcg | ctgagatagg | 4500 |
| | egecteacty | attaagcatt | qqtaactqtc | agaccaagtt | tactcatata | tactttacat | 1560 |
| | cyatttaaaa | CLICALLILL | aatttaaaag | gatetaggtg | aadatccttt | ttgataatct | 1620 |
| | catyaccaaa | atcccttaac | graagtttc | Qttccactga - | acatcagacc | ccutagaaa | 1690 |
| C C | gaccaaagga | LCLLCLLgag | atccttttt | TCtgcgcgta - | atctactact | tacasacasa | 4740 |
| 60 | addaccaccg | ctaccagcgg | taatttattt | QCCqqatcaa | gagetaceaa | ctctttttcc | 4000 |
| | gaaggtaact | ggcttcagca | gagegeagat | 3syucoud | atacttata | tataaaa-t- | 4000 |
| | gttaggccac | cacttcaaga | actototage | acconct | tageter-t- | tacks-t | 4000 |
| | gttaccanto | actactacce | ataacastaa | aboguctaca | Lacotogoto | cyclaatect | 4920 |
| | atauttacca | gctgctgcca | acceptance | gregtgtett | accgggttgg | actcaagacg | 4980 |
| 65 | cttocaccy | gataaggcgc | aycygtcggg | ccgaacgggg | ggttcgtgca | cacagcccag | 5040 |
| 55 | uccygaydya | acqacctaca | ccgaactgag | atacctacad | catagactat | dadaaadcac | 5100 |
| | Cacyclice | qaaqqqaqaa | aggeggaeag | Otateconta | ancaacaaaa | transacana | 5160 |
| | ayagcgcacg | agggagcttc | cagggggaaa | cgcctggtat | ctttatagtc | ctgtcgggtt | 5220 |
| | | | | • | . | | - |

```
8
      tegecacete tgaettgage gtegattttt gtgatgeteg teagggggge ggageetatg 5280
      gaaaaacgcc agcaacgcgg cettttacg gttcctggcc ttttgctggc cttttgctca 5340 catgttcttt cctgcgttat cccctgattc tgtggataac cgtattaccg cetttgagtg 5400
      agetgatace getegeegea geegaacgae egagegeage gagteagtga gegaggaage 5460
      ggaagagege ccaataegea aacegeetet eccegegegt tggeegatte attaatgeag 5520
      ttagctcact cattaggcac cccaggcttt acactttatg cttccggctc gtatgttgtg 5640
      tggaattgtg agcggataac aatttcacac aggaaacagc tatgaccatg attacgccaa 5700
      gctagcccgg gctagcttgc atg
 10
      <210> 13
      <211> 4960
      <212> DNA
 15
      <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: vector
            pCMV-Cre
 20
      <400> 13
      aaacagteeg atgtaeggge cagatataeg egttgacatt gattattgae tagttattaa 60
      tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg cgttacataa 120
      cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt gacgtcaata 180
25
      atgacgtatg tteccatagt aaegecaata gggactitee attgacgtea atgggtggae 240
      tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc aagtacgccc 300
      cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta catgacctta 360
     tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac catggtgatg 420
      cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg atttccaagt 480
30
     ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg ggactttcca 540
     aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt acggtgggag 600
     gtctatataa gcagagctct ctggctaact agagaaccca ctgcttactg gcttatcgaa 660
     attaatacga etcaetatag ggagacecaa getgaeteta gaettaatta agegttgggg 720
     tgagtactcc ctctcaaaag cgggcatgac ttctgcgcta agattgtcag tttccaaaaa 780
35
     cgaggaggat ttgatattca cctggcccgc ggtgatgcct ttgagggtgg ccgcgtccat 840
     ctggtcagaa aagacaatct ttttgttgtc aagcttgagg tgtggcaggc ttgagatctg 900
     gocatacact tgagtgacat tgacatccac tttgcctttc tctccacagg tgtccactcc 960
     cagggcggcc tcgaccatgc ccaagaagaa gaggaaggtg tccaatttac tgaccgtaca 1020
     ccaaaatttg cctgcattac cggtcgatgc aacgagtgat gaggttcgca agaacctgat 1080
40
     ggacatgttc agggatcgcc aggcgttttc tgagcatacc tggaaaatgc ttctgtccgt 1140
     ttgccggtcg tgggcggcat ggtgcaagtt gaataaccgg aaatggtttc ccgcagaacc 1200
     tgaagatgtt cgcgattatc ttctatatct tcaggcgcgc ggtctggcag taaaaactat 1260
     ccagcaacat ttgggccagc taaacatgct tcatcgtcgg tccgggctgc cacgaccaag 1320
     tgacagcaat gctgtttcac tggttatgcg gcggatccga aaagaaaacg ttgatgccgg 1380
45
     tgaacgtgca aaacaggete tagegttega aegcaetgat ttegaecagg ttegtteaet 1440
     catggaaaat agcgateget gccaggatat acgtaatetg gcatttetgg ggattgetta 1500
     taacaccctg ttacgtatag ccgaaattgc caggatcagg gttaaagata tctcacgtac 1560
     tgacggtggg agaatgttaa tccatattgg cagaacgaaa acgctggtta gcaccgcagg 1620
     tgtagagaag gcacttagcc tgggggtaac taaactggtc gagcgatgga tttccgtctc 1680
50
     tggtgtaget gatgateega ataactaeet gttttgeegg gteagaaaa atggtgttge 1740
     egegecatet gecaccagee agetateaac tegegecetg gaagggattt ttgaagcaac 1800
     tcatcgattg atttacggcg ctaaggatga ctctggtcag agatacctgg cctggtctgg 1860
     acacagtgcc cgtgtcggag ccgcgcgaga tatggcccgc gctggagttt caataccgga 1920
     gatcatgcaa gctggtggct ggaccaatgt aaatattgtc atgaactata tccgtaacct 1980
55
     ggatagtgaa acaggggcaa tggtgcgcct gctggaagat ggcgattagc cattaacgcg 2040
     taaatgattg cagatccact agttctaggg ccgcgtcgac ctcgagatcc aggcgcggat 2100
     caataaaaga tcattatttt caatagatct gtgtgttggt tttttgtgtg ccttggggga 2160
     gggggaggcc agaatgaggc gcggccaagg gggagggga ggccagaatg accttggggg 2220
     agggggagge cagaatgace ttgggggagg gggaggccag aatgaggege geeceegggt 2280 acegageteg aatteaetgg eegtegttt acaacgtegt gaetgggaaa aceetggegt 2340
60
     tacccaactt aatcgccttg cagcacatcc ccctttcgcc agctggcgta atagcgaaga 2400
     ggcccgcacc gatcgccctt cccaacagtt gcgcagcctg aatggcgaat ggcgcctgat 2460
     goggtatttt etecttaege atetgtgegg tattteacae egeatatggt geacteteag 2520
     tacaatetge tetgatgeeg catagttaag eeageeeega caeeegeeaa caeeegetga 2580 egegeeetga egggettgte tgeteeegge ateegettae agacaagetg tgaeegtete 2640
65
     cgggagctgc atgtgtcaga ggttttcacc gtcatcaccg aaacgcgcga gacgaaaggg 2700
     cctcgtgata cgcctatttt tataggttaa tgtcatgata ataatggttt cttagacgtc 2760
```

```
aggtggcact tttcggggaa atgtgcgcgg aacccctatt tgtttatttt tctaaataca 2820
      ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat aatattgaaa 2880
      aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt ttgcggcatt 2940
      ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg ctgaagatca 3000
      gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag 3060
      ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagttctgc tatgtggcgc 3120
      ggtattatec egtattgaeg eegggeaaga geaacteggt egeegeatae actattetea 3180
      gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg gcatgacagt 3240
      aagagaatta tgcagtgctg ccataaccat gagtgataac actgcggcca acttacttct 3300
 10
      gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg gggatcatgt 3360
      aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg acgagcgtga 3420
      caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg gcgaactact 3480 tactctagct tcccggcaac aattaataga ctggatggag gcggataaag ttgcaggacc 3540
      acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg gagccggtga 3600
15
      gcgtgggtct cgcggtatca ttgcagcact gggggccagat ggtaagccct cccgtatcgt 3660
      agttatetae acgaegggga gteaggeaae tatggatgaa egaaatagae agategetga 3720
      gataggtgcc tcactgatta agcattggta actgtcagac caagtttact catatatact 3780
      ttagattgat ttaaaacttc attttaatt taaaaggatc taggtgaaga tcctttttga 3840
      taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt cagaccccgt 3900
20
      agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct gctgcttgca 3960
      aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct 4020
     ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc ttctagtgta 4080 gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc tcgctctgct 4140
      aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg ggttggactc 4200
      aagacgatag ttaccggata aggcgcagcg gtcgggctga acggggggtt cgtgcacaca 4260
     gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg agctatgaga 4320
     aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg 4380
     aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggtatcttt atagtcctgt 4440
     cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag gggggcggag 4500
30
     cctatggaaa aacgccagca acgcggcctt tttacggttc ctggcctttt gctggccttt 4560
     tgctcacatg ttctttcctg cgttatcccc tgattctgtg gataaccgta ttaccgcctt 4620
     tgagtgaget gatacegete geegeageeg aacgacegag egeagegagt cagtgagega 4680 ggaageggaa gagegeecaa tacgcaaace geeteteeee gegegttgge egatteatta 4740
     atgcagetgg cacgacaggt ttecegactg gaaageggge agtgagegea acgeaattaa 4800
35
     tgtgagttag ctcactcatt aggcacccca ggctttacac tttatgcttc cggctcgtat 4860
     gttgtgtgga attgtgagcg gataacaatt tcacacagga aacagctatg accatgatta 4920
     cgccaagcta gcccgggcta gcttgcatgc ctgcaggttt
40
     <210> 14
     <211> 3858
     <212> DNA
     <213> Artificial Sequence
45
     <220>
     <223> Description of Artificial Sequence: vector pRK50
     aaacagtccg atgtacgggc cagatatacg cgttgacatt gattattgac tagttattaa 60
50
     tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg cgttacataa 120
     cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt gacgtcaata 180
     atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca atgggtggac 240
     tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc aagtacgccc 300
     cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta catgacctta 360
55
     tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac catggtgatg 420
     cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg atttccaagt 480
     ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg ggactttcca 540
     aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt acggtgggag 600
     gtctatataa gcagagctct ctggctaact agagaaccca ctgcttactg gcttatcgaa 660
     attaatacga ctcactatag ggagacccaa gctgactcta gacttaatta agcgttgggg 720
     tgagtactcc ctctcaaaag cgggcatgac ttctgcgcta agattgtcag tttccaaaaa 780
     cgaggaggat ttgatattca cctggcccgc ggtgatgcct ttgagggtgg ccgcgtccat 840
     ctggtcagaa aagacaatct ttttgttgtc aagcttgagg tgtggcaggc ttgagatctg 900
     gecatacact tgagtgacat tgacatecae tttgcctttc tctccacagg tgtccactcc 960
65
     cagggcggcc gcgtcgacct cgagatccag gcgcggatca ataaaagatc attatttca 1020
```

atagatetgt gtgttggttt tttgtgtgee ttgggggagg gggaggeeag aatgaggege 1080 ggeeaagggg gaggggagg eeagaatgae ettgggggag ggggaggeea gaatgaeett 1140

```
gggggaggg gaggccagaa tgaggcgcgc ccccgggtac cgagctcgaa ttcactggcc 1200
      gtcgttttac aacgtcgtga ctgggaaaac cctggcgtta cccaacttaa tcgccttgca 1260
      gcacatcece etttegeeag etggegtaat agegaagagg eeegeacega tegeeettee 1320
      caacagttgc gcagcctgaa tggcgaatgg cgcctgatgc ggtattttct ccttacgcat 1380
      ctgtgcggta tttcacaccg catatggtgc actctcagta caatctgctc tgatgccgca 1440
      tagttaagcc agccccgaca cccgccaaca cccgctgacg cgccctgacg ggcttgtctg 1500 ctcccggcat ccgcttacag acaagctgtg accgtctccg ggagctgcat gtgtcagagg 1560 ttttcaccgt catcaccgaa acgcgcgaga cgaaagggcc tcgtgatacg cctatttta 1620 taggttaatg tcatgataat aatggttct tagacgtcag gtggcacttt tcggggaaat 1680
 10
      gtgcgcggaa cccctatttg tttattttc taaatacatt caaatatgta tccgctcatg 1740
      agacaataac cctgataaat gcttcaataa tattgaaaaa ggaagagtat gagtattcaa 1800
      cattlecgtg tegecettat tecettttt geggeatttt geetteetgt ttttgeteae 1860
      ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt tgggtgcacg agtgggttac 1920
      atcgaactgg atctcaacag cggtaagatc cttgagagtt ttcgccccga agaacgtttt 1980
      ccaatgatga gcacttttaa agttctgcta tgtggcgcgg tattatcccg tattgacgcc 2040
      gggcaagagc aactcggtcg ccgcatacac tattctcaga atgacttggt tgagtactca 2100
      ccagteacag aaaagcatet tacggatgge atgacagtaa gagaattatg cagtgetgee 2160
      ataaccatga gtgataacac tgcggccaac ttacttctga caacgatcgg aggaccgaag 2220
      gagetaaceg etttttgca caacatgggg gateatgtaa etegeettga tegttgggaa 2280
      ccggagetga atgaagecat accaaacgae gagegtgaca ccaegatgee tgtageaatg 2340
20
      gcaacaacgt tgcgcaaact attaactggc gaactactta ctctagcttc ccggcaacaa 2400
      ttaatagact ggatggaggc ggataaagtt gcaggaccac ttctgcgctc ggcccttccg 2460
      gctggctggt ttattgctga taaatctgga gccggtgagc gtgggtctcg cggtatcatt 2520
      gcagcactgg ggccagatgg taagccctcc cgtatcgtag ttatctacac gacggggagt 2580
25
      caggcaacta tggatgaacg aaatagacag atcgctgaga taggtgcctc actgattaag 2640
      cattggtaac tgtcagacca agtttactca tatatacttt agattgattt aaaacttcat 2700
      ttttäättta aaaggateta ggtgaagate etttttgata ateteatgae caaaateeet 2760
      taacgtgagt tttcgttcca ctgagcgtca gaccccgtag aaaagatcaa aggatcttct 2820
      tgagatectt tttttetgeg egtaatetge tgettgeaaa caaaaaaace accgetacea 2880
30
      geggtggttt gtttgeegga teaagageta ceaactettt tteegaaggt aactggette 2940
      agcagagege agataceaaa taetgteett etagtgtage egtagttagg ceaceaette 3000
      aagaactetg tagcaccgcc tacatacctc getetgetaa teetgttacc agtggetget 3060
      gccagtggcg ataagtcgtg tcttaccggg ttggactcaa gacgatagtt accggataag 3120
     gcgcagcggt cgggctgaac ggggggttcg tgcacacagc ccagcttgga gcgaacgacc 3180
     tacaccgaac tgagatacct acagcgtgag ctatgagaaa gcgccacgct teeegaaggg 3240
     agaaaggegg acaggtatee ggtaagegge agggteggaa caggagageg caegagggag 3300
     cttccagggg gaaacgcctg gtatctttat agtcctgtcg ggtttcgcca cctctgactt 3360 gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc tatggaaaaa cgccagcaac 3420 gcggcctttt tacggttcct ggccttttgc tggccttttg ctcacatgtt ctttcctgcg 3480
40
     ttatcccctg attotgtgga taaccgtatt accgcctttg agtgagctga taccgctcgc 3540
     cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg aagcggaaga gcgcccaata 3600
     cgcaaaccgc ctctccccgc gcgttggccg attcattaat gcagctggca cgacaggttt 3660
     cccgactgga aagcgggcag tgagcgcaac gcaattaatg tgagttagct cactcattag 3720
     gcaccccagg ctttacactt tatgcttccg gctcgtatgt tgtgtggaat tgtgagcgga 3780
45
     taacaattic acacaggaaa cagctatgac catgattacg ccaagctage ccgggctage 3840
     ttgcatgcct gcaggttt
     <210> 15
50
     <211> 6257
     <212> DNA
     <213> Artificial Sequence
     <220>
55
     <223> Description of Artificial Sequence: vector
            pRK64(deltaCre)
     <400> 15
     cgtcatcacc gaaacgcgcg aggcagctgt ggaatgtgtg tcagttaggg tgtggaaagt 60
60
     ccccaggete eccageagge agaagtatge aaagcatgea teteaattag teageaacea 120
     ggctccccag caggcagaag tgtgcaaagc atgcatctca attagtcagc aaccatagtc 180
     cogeccetaa eteogeceat ecegeceeta acteegecea gtteegecea tteteegece 240
     catggctgac taatttttt tatttatgca gaggccgagg ccgcctcggc ctaggaacag 300
     togacgacac tgcagagacc tacttcacta acaaccggta cagttcgtgg accagatggg 360
65
     tgaggtggag tacgcgccg gggagcccaa gggcacgccc tggcacccgc accgcggctt 420
     cgagaccgtc acgaatagat ccataacttc gtatagcata cattatacga agttataccg 480
     ggccaccatg gtcgcgagta gcttggcact ggccgtcgtt ttacaacgtc gtgactggga 540
```

| | aaaccctggc | gttacccaac | ttaatcgcct | tgcagcacat | cccctttcq | ccagctggcg | 600 |
|----|------------|-------------|-------------|------------|-------------|--------------------------|------|
| | taatagcgaa | gaggcccgca | ccgatcgccc | ttcccaacag | ttgcgcagcc | tgaatgggga | 660 |
| | atggcgcttt | gcctqgtttc | cggcaccaga | agcggtgccg | gaaagetgge | tagaatacaa | 720 |
| _ | tetteetgag | gccgatactg | tcatcatccc | ctcaaactgg | cagatgcacg | attacaatac | 780 |
| 5 | gcccatctac | accaacgtaa | cctatcccat | tacootcaat | ccaccattta | ttcccacaa | 840 |
| | gaatccgacg | ggttgttact | coctcacatt | taatottoat | gaaagctggc | tacaggaagg | 900 |
| | ccagacgcga | attatttttg | atggcgttaa | ctcagcattt | catctgtggt | gcaacgggcg | 960 |
| | ctgggtcggt | tacggccagg | acagtcgttt | accatctasa | tttgacctga | gcgcattttt | 1020 |
| | acgcgccgga | gaaaaccgcc | tcacaataat | gatactacat | tagaccega | gcagttatct | 1020 |
| 10 | ggaagatcag | gatatotogo | agatgaggg | cattttccct | asateteat | tgctgcataa | 1140 |
| | accoactaca | caaatcagcg | atttccatot | taccactcac | tttaataata | atttcagccg | 1200 |
| | cactatacta | gaggetgaag | ttcagatgtg | caacaaatta | catalctaca | tacgggtaac | 1260 |
| | agtttcttta | tagcagggtg | aaacgcagg | caccaacaac | accordatt | tcggcggtga | 1200 |
| | aattatcoat | gagcatagta | attataccaa | tcacatcaca | ctacatataa | acgtcgaaaa | 1320 |
| 15 | cccgaaactg | fagaacacca | aaatcccga | tetetateat | acastastta | actgcacac | 1300 |
| | caccascaac | acoctoatto | aagcagaagc | ctacaatata | geggeggeeg | aggtgcggat | 1500 |
| | tgaaaatggt | ctactactac | tgaacggcaa | accattacta | attenagge | ttaaccgtca | 1560 |
| | cgagcatcat | cctctgcatg | gtcaggtcat | googeegeeg | accestaggeg | aggatatcct | 1620 |
| | gctgatgaag | cagaacaact | ttaacgccgt | acactattca | cattatecea | accatecget | 1600 |
| 20 | gtggtacacg | ctgtgcgacc | gctacggcct | gtatgtggtg | gatgaagcca | atattgaaac | 1740 |
| | ccacggcatg | gtgccaatga | atcotctoac | coatoatcco | cactaactac | cggcgatgag | 1800 |
| | cgaacgcgta | acqcqaatqq | tgcagcgcga | tcgtaatcac | ccaaatataa | tcatctggtc | 1860 |
| | gctggggaat | gaatcaggcc | acggcgctaa | tcacgacgcg | ctgtatcgct | ggatcaaatc | 1920 |
| | tgtcgatcct | tecegeeegg | tgcagtatga | addcddcdda | accascacca | cggccaccga | 1980 |
| 25 | tattatttgc | ccgatgtacg | cacacataa | tgaagaccag | cccttcccaa | ctgtgccgaa | 2040 |
| | atggtccatc | aaaaaatggc | tttcgctacc | tggagagacg | cacccactaa | tcctttgcga | 2100 |
| | atacgcccac | gcgatgggta | acagicttgg | caatttcact | aaatactggc | aggcgtttcg | 2160 |
| | tcagtatccc | cgtttacagg | gcggcttcgt | ctgggactgg | gtggatcagt | cgctgattaa | 2220 |
| | atatgatgaa | aacggcaacc | cgtggtcggc | ttacqqcqqt | gattttggcg | atacoccoaa | 2280 |
| 30 | cgatcgccag | ttctgtatga | acggtctggt | ctttqccqac | cqcacqccqc | atccagcgct | 2340 |
| | gacggaagca | aaacaccagc | agcagttttt | ccagttccgt | ttatccgggc | aaaccatcga | 2400 |
| | agtgaccagc | gaatacctgt | tccgtcatag | cgataacgag | ctcctgcact | ggatggtggc | 2460 |
| | gctggatggt | aagccgctgg | caagcggtga | agtgcctctg | gatgtcgctc | cacaaggtaa | 2520 |
| | acagttgatt | gaactgcctg | aactaccgca | qccqqaqaqc | gccgggcaac | tetageteac | 2580 |
| 35 | agtacgcgta | gtgcaaccga | acgcgaccgc | atggtcagaa | gccgggcaca | tcagcgcctg | 2640 |
| | gcagcagtgg | cgtctggcgg | aaaacctcag | tgtgacgctc | cccaccacat | cccacaccat | 2700 |
| | cccgcatctg | accaccageg | aaatggattt | ttgcatcgag | ctgggtaata | agcattagca | 2760 |
| | atttaaccgc | cagtcaggct | ttctttcaca | gatgtggatt | ggcgataaaa | aacaactoct | 2820 |
| 40 | gacgccgctg | cgcgatcagt | tcacccgtqc | accgctggat | aacqacatto | gcgtaagtga | 2880 |
| 40 | agcgacccgc | attgacccta | acgcctgggt | cgaacgctgg | aaggcggcgg | gccattacca | 2940 |
| | ggccgaagca | gcgttgttgc | agtgcacggc | agatacactt | gctgatgcgg | toctoattac | 3000 |
| | gaccgctcac | gcgtggcagc | atcaggggaa | aaccttattt | atcagccgga | aaacctacco | 3060 |
| | gattgatggt | agtggtcaaa | tggcgattac | cottoatott | gaagtggcga | gcgatacacc | 3120 |
| 45 | gcatccggcg | cggattggcc | tgaactgcca | gctggcgcag | gtagcagagc | gggtaaactg | 3180 |
| 45 | gctcggatta | gggccgcaag | aaaactatcc | cgaccgcctt | actoccocct | gttttgaccg | 3240 |
| | ctgggatctg | ccattgtcag | acatgtatac | cccgtacgtc | ttcccgagcg | aaaacggtct | 3300 |
| | gcgctgcggg | acgcgcgaat | tgaattatgg | cccacaccag | tggcgcggcg | acttccagtt | 3360 |
| | caacatcagc | cgctacagtc | aacagcaact | gatggaaacc | agccatcgcc | atctgctgca | 3420 |
| 50 | cycggaagaa | ggcacatggc | tgaatatcga | cggtttccat | atggggattg | gtggcgacga | 3480 |
| 50 | ctcctggage | ccgtcagtat | cggcggaatt | ccagctgagc | gccggtcgct | accattacca | 3540 |
| | greggeergg | tgtcaaaaat | aataataacc | gggcaggggg | gatctttgtg | aaggaacctt | 3600 |
| | acticigegg | tgtgacataa | ttggacaaac | tacctacaga | gatttaaagc | tctaaggtaa | 3660 |
| | acacacact | tttaagtgta | taatgtgtta | aactactgat | tctaattgtt | tgtgtatttt | 3720 |
| 55 | ayattccaac | ctatggaact | gatgaatggg | agcagtggtg | gaatgccaga | tccagacatg | 3780 |
| 33 | ataagataca | ttgatgagtt | tggacaaacc | acaactagaa | tgcagtgaaa | aaaatgcttt | 3840 |
| | atttgtgaaa | tttgtgatgc | tattgcttta | tttgtaacca | ttataagctg | caataaacaa | 3900 |
| | gitaacaaca | acaattgcat | tcattttatg | tttcaggttc | agggggaggt | gtgggaggtt | 3960 |
| | ccctaaagca | agtaaaacct | ctacaaatgt | ggtatggctg | attatgatct | gcggccgcag | 4020 |
| 60 | ggeetegtga | tacgcctatt | tttataggtt | aatgtcatga | taataatggt | ttcttagacg | 4080 |
| 00 | cattorate | cttttcgggg | aaatgtgcgc | ggaaccccta | tttgtttatt | tttctaaata | 4140 |
| | accudadta | cgtatccgct | tacyagacaa | taaccctgat | aaatgcttca | ataatattga | 4200 |
| | ttttaastt | grargagtat | toaccatttc | cgtgtcgccc | ttattccctt | ttttgcggca | 4260 |
| | Canttenat~ | orderetted. | ttacatcagaa | acgctggtga | aagtaaaaga | tgctgaagat | 4320 |
| 65 | agttttccc | cacgagtggg | ttttccs. | ctggatctca | acagcggtaa | gatccttgag | 4380 |
| 55 | acaut=++=+ | cogadyddog | caccacacaca | atgagcactt | ctaaagttct | gctatgtggc | 4440 |
| | Cadaatdact | trattaeate | ctcaccacta | yagcaactcg | gregeegeat | acactattct tggcatgaca | 4500 |
| | gualyact | cygucyayua | Judged | acagaaaagc | accitacgga | Lygcatgaca | 4260 |

```
gtaagagaat tatgcagtgc tgccataacc atgagtgata acactgcggc caacttactt 4620
     ctgacaacga tcggaggacc gaaggagcta accgcttttt tgcacaacat gggggatcat 4680
     gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag ccataccaaa cgacgagcgt 4740
     gacaccacga tgcctgtagc aatggcaaca acgttgcgca aactattaac tggcgaacta 4800
     cttactctag cttcccggca acaattaata gactggatgg aggcggataa agttgcagga 4860
     ccacttctgc gctcggccct tccggctggc tggtttattg ctgataaatc tggagccggt 4920
     gagcgtgggt ctcgcggtat cattgcagca ctggggccag atggtaagcc ctcccgtatc 4980
     gtagttatct acacgacggg gagtcaggca actatggatg aacgaaatag acagatcgct 5040
     gagataggtg cctcactgat taagcattgg taactgtcag accaagttta ctcatatata 5100
10
     ctttagattg atttaaaact tcatttttaa tttaaaagga tctaggtgaa gatccttttt 5160
     gataatetea tgaccaaaat ceettaaegt gagttttegt teeactgage gteagaceee 5220
     gtagaaaaga tcaaaggatc ttcttgagat ccttttttc tgcgcgtaat ctgctgcttg 5280
     caaacaaaaa aaccaccgct accagcggtg gtttgtttgc cggatcaaga gctaccaact 5340
     ctttttccga aggtaactgg cttcagcaga gcgcagatac caaatactgt ccttctagtg 5400
     tagccgtagt taggccacca cttcaagaac tctgtagcac cgcctacata cctcgctctg 5460
     ctaatcotgt taccagtggc tgctgccagt ggcgataagt cgtgtcttac cgggttggac 5520
     tcaagacgat agttaccgga taaggcgcag cggtcgggct gaacgggggg ttcgtgcaca 5580
     cageceaget tggagegaac gacetacace gaactgagat acetacageg tgagetatga 5640
     gaaagegeea egetteeega agggagaaag geggacaggt ateeggtaag eggeagggte 5700
20
     ggaacaggag agcgcacgag ggagcttcca ggggggaaacg cctggtatct ttatagtcct 5760
     gtcgggtttc gccacctctg acttgagcgt cgatttttgt gatgctcgtc aggggggcgg 5820
     agectatgga aaaaegecag caaegeggee titttaeggt teetggeett tigetggeet 5880
     tttgctcaca tgttctttcc tgcgttatcc cctgattctg tggataaccg tattaccgcc 5940
     tttgagtgag ctgataccgc tcgccgcagc cgaacgaccg agcgcagcga gtcagtgagc 6000
    gaggaagcgg aagagcgcc aatacgcaaa ccgcctctcc ccgcgcgttg gccgattcat 6060 taatgcagct ggcacgacag gtttcccgac tggaaagcgg gcagtgagcg caacgcaatt 6120
     aatgtgagtt agctcactca ttaggcaccc caggctttac actttatgct tccggctcgt 6180
     atgttgtgtg gaattgtgag cggataacaa tttcacacag gaaacagcta tgaccatgat 6240
     tacgccaagc tggcgcg
30
     <210> 16
     <211> 6252
     <212> DNA
35
     <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: vector
           pRK64 (deltaInt)
40
     <400> 16
     cgtcatcacc gaaacgcgcg aggcagctgt ggaatgtgtg tcagttaggg tgtggaaagt 60
     ecceaggete eccageagge agaagtatge aaageatgea teteaattag teageaacea 120
     ggctccccag caggcagaag tgtgcaaagc atgcatctca attagtcagc aaccatagtc 180
45
     cogococtaa ctoogococta actoogocoa gttoogocoa ttotoogoco 240
     catggctgac taatttttt tatttatgca gaggccgagg ccgcctcggc ctaggaacag 300
     tegacgaeae tgeagagaee taetteaeta acaaceggta cagttegtgg accagatggg 360
     tgaggtggag tacgcgcccg gggagcccaa aggttacccc agttggggca ctactcccga 420
     aaaccgcttc tggatccata acttcgtata gcatacatta tacgaagtta taccgggcca 480
50
     ccatggtcgc gagtagettg geactggccg tegttttaca acgtcgtgac tgggaaaacc 540
     ctggcgttac ccaacttaat cgccttgcag cacatccccc tttcgccagc tggcgtaata 600
     gcgaagaggc ccgcaccgat cgcccttccc aacagttgcg cagcctgaat ggcgaatggc 660
     getttgeetg gttteeggea ecagaagegg tgeeggaaag etggetggag tgegatette 720 etgaggeega taetgtegte gteeceteaa aetggeagat geaeggttae gatgegeeca 780
55
     tctacaccaa cgtaacctat cccattacgg tcaatccgcc gtttgttccc acggagaatc 840
     cgacgggttg ttactcgctc acatttaatg ttgatgaaag ctggctacag gaaggccaga 900
     cgcgaattat ttttgatggc gttaactcgg cgtttcatct gtggtgcaac gggcgctggg 960
     tcggttacgg ccaggacagt cgtttgccgt ctgaatttga cctgagcgca tttttacgcg 1020
     ccggagaaaa ccgcctcgcg gtgatggtgc tgcgttggag tgacggcagt tatctggaag 1080 atcaggatat gtggcggatg agcggcattt tccgtgacgt ctcgttgctg cataaaccga 1140
60
     ctacacaaat cagcgatttc catgttgcca ctcgctttaa tgatgatttc agccgcgctg 1200
     tactggaggc tgaagttcag atgtgcggcg agttgcgtga ctacctacgg gtaacagttt 1260
     ctttatggca gggtgaaacg caggtcgcca gcggcaccgc gcctttcggc ggtgaaatta 1320
     togatgageg tggtggttat googatogeg toacactacg totgaacgto gaaaaccoga 1380
     aactgtggag cgccgaaatc ccgaatctct atcgtgcggt ggttgaactg cacaccgccg 1440
     acggcacget gattgaagca gaagcetgeg atgteggttt cegegaggtg eggattgaaa 1500
     atggtetget getgetgaac ggeaageegt tgetgatteg aggegttaac egteacgage 1560
```

PCT/EP01/12975

•

| | | | | 13 | | | |
|------------|---------------------------|--------------------|--------------|---|--|-------------------|---------|
| | atcatcctct | gcatggtcag | gtcatggatg | g agcagacgat | ggtgcaggat | atcetectea | 1620 |
| | • Jung cugaa | Laatttaat | - accaracact | . <i>attearatta</i> | + + + + + + + + + + + + + + + + + + + | | |
| | googeg | , cyaccyctat | - quectatata | , taaraaaraa | ACCCAA+ | ~~~~~~ | 1740 |
| 5 | 3 | - uucguuccy c | Ctuationali | i arccococro | I GCTACCGGGG | +~-~ | 1000 |
| J | gogcaacgcg | aatyytycay | COCCATCATA | atrarrran | , tatastasta | | 1000 |
| | gg waa a gaac c | aggccacggc | quiaatcacc | i acococtota | trartareta | 22246464 | 1000 |
| | | - occay cycay | Latuaduucc | , oroganeroa | | ~ | 1000 |
| | vg-ccgac | gracycycyc | ququatqaac | Laccadecett | cccaactata | 2222224 | 2040 |
| 10 | oou ccaaaaa | aryychticg | ctacctggad | i agacgcgccc | : actastact | taaaaataaa | 2100 |
| 10 | oooucgcgat | yyyraacayt | CLLaacaart | teochaaata | ctaacaaaaa | ++++ | 01.00 |
| | accongic | acayyycyyc | rrcarcraad | r actodotoda | tragtracta | a++ a a a + a + a | 2222 |
| | | Juacecycy | LCGGCLLaco | l dcddraattt | taacastsaa | 202222 | 2222 |
| | goodgeceeg | Latyaatyyt | CLUGECTER | CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | accacataca | ~~~~ | 0040 |
| 1.5 | augcadaaca | ccaycaycay | LLLLLCCagr | tecentitate | COCCCSSSS | 2+cc=+ | 2400 |
| 15 | oougoguuca | CCLULLCCULL | catagegata | . accaactcct | acactacata | ~+~~~~~+ | 0460 |
| | a a g g a a a g a c | yctyycaayc | quidaadedc | CTCTGGATGT | CCC CCC CCC | aat saaaa | 2524 |
| | | goocyaacta | CCUCADECO | anancecca | ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ | ~ | 0 - 0 0 |
| | 3-3-49-9-4 | accyaacycy | accucator | Cagaageegg | " deacateage | ~~~+~~~~ | 2000 |
| | - 3 - 3 3 - 3 | ggoggaaaac | CLCaululud | . cacreecer | COCOTCCC | ACCOST COOK | 2700 |
| 20 | acceded | caycyaaaty | uattitioca | - TCGAGCTGGG | taataadoot | +~~~~~ | 2760 |
| | arrage cage c | aggette | LCacadaror | ananttann | + | ~ | 0000 |
| | -55-09-09u | | CULUCACION | ranaraacaa | 02ttacaata | ~ ~ + ~ | 0000 |
| | | CCCCaacac | Luuulcaaac | actadaadac | aacaaaaaa | + | 2040 |
| 0.5 | | gergeagege | acuucauala | Cacttoctoa | TACAATAATA | 2++22 | 2000 |
| 25 | gugug | geageateag | uuudadacct | Lattratead | CCCCCCCCCC | + | 3000 |
| | arggragegg | ccaaatyycy | attaccotto | atorroaact | dacasacast. | 20200001 | 2120 |
| | | Lygoutuaat | Luccaderaa | CUCAGGEAGC | adaddddata | | 2100 |
| | gaccagggcc | y vaayaaaac | Latececaace | accettactac | ~~~~++++ | | ~~ ~ ~ |
| 30 | | gucagacatq | Latacccccc | acorerrece | Tarres and a | antatassas | 2200 |
| 30 | J- JJJ~~909 | - Guuttyaat | Latuacccac | accagrages | COOCCAPCEEC | anatten | 2260 |
| | a a a g a a g a a g a a g | cayttaatay | Caactgatgg | aaaccaccca | tegeesteta | atanaa | 2420 |
| | | ucuucuaat | accuacoone | TCCSTSTGGG | 7377777777 | | 2 4 6 6 |
| | ggagaaaagaa | agrarcyggg | gaattccagc | Lusucaccaa | teactacost | + | 2540 |
| 35 | 555000 | uuuu caa caa | Laaccuuuca | udddddarcr | ttatassaas | 200112011 | 3600 |
| | | cataattyya | Cadactacct | acadadattt | aaadetetaa | aataaatat. | 3660 |
| | uuuccccaa | gryraraary | LULTABACTA | CEGATECTAA | ++~++~+~+ | -+++ | 2700 |
| | atacattgat | gaactyatya | argggagcag | tggtggaatg | ccagatccag | acatgataag | 3780 |
| | tgaaatttgt | gageeeggae | ctttattat | tagaatgcag | tgaaaaaaat | gctttatttg | 3840 |
| 40 | caacaacaat | tocattcatt | ttatettte | aaccattata | agctgcaata | aacaagttaa | 3900 |
| | aagcaagtaa | aacctctaca | aatataatat | ggttcagggg | gaggtgtggg gatctgcggc | aggttttta | 3960 |
| | cgtgatacgc | ctatttttat | addteatat | ggccgaccac | atggtttctt | cgcagggcct | 4020 |
| | | | | | | | |
| | addiatytat | Coucidata | gacaaraacc | ctaataaata | C++C+ | | 4000 |
| 45 | | | | | | | |
| | Joe Locky Lt | LLLGGLCACC | cadaaacdct | adtassadts. | 2220240040 | | 4200 |
| | | | | | | | |
| | - Togccccyaa | gaacquitte | caatdatdad | Cacttttaaa' | " | ~+~~~~~ | A A A A |
| 50 | cocgc | uccyacyccy | uucaauauca | acteggtege | CCCatacact | 2++2+22-2 | 4 E O O |
| 50 | | ququucccac | COULCACAGA | aaadcatet | ~~~~~~~~ | | 4566 |
| | ~g~accacgc | agrycrycca | Laaccatgag | rgaraacact | acaacaaaat | ++ | 4.000 |
| | | | | | | | |
| | | | CUUAOCT OAA | Traarreats | ~~~~~~~ | | 4740 |
| 55 | | g cag caa caa | Caacaacur | CCCC222Ct2 | ++ | | 4000 |
| 55 | | | | | | | |
| | | 900000000 | CLUUCH.CRIT | Fattacras+ | | | 4000 |
| | | | | | | | |
| | | | | | | | |
| <i>c</i> n | | | | | | | |
| 60 | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| 65 | | | | | | | |
| 65 | | | | | | | |
| | | | | | | | |
| | acgatagtta | ccggataagg | cgcagcggtc | gggctgaacq | gggggttcgt | gcacacagee | 5580 |
| | | | | - | | | |

```
cagcttggag cgaacgacct acaccgaact gagataccta cagcgtgagc tatgagaaag 5640
      cgccacgett cccgaaggga gaaaggcgga caggtatccg gtaagcggca gggtcggaac 5700
      aggagagege acgagggage ttecaggggg aaacgcetgg tatettata gteetgtegg 5760 gtttegecae etetgaettg agegtegatt tttgtgatge tegteagggg ggeggageet 5820
      atggaaaaac gccagcaacg cggccttttt acggttcctg gccttttgct ggccttttgc 5880
      tcacatgttc tttcctgcgt tatcccctga ttctgtggat aaccgtatta ccgcctttga 5940
      gtgagctgat accgctcgcc gcagccgaac gaccgagcgc agcgagtcag tgagcgagga 6000
      ageggaagag egeceaatae geaaacegee teteceegeg egttggeega tteattaatg 6060
      cagctggcac gacaggtttc ccgactggaa agcgggcagt gagcgcaacg caattaatgt 6120
 10
      gagttagete acteattagg caccecagge titacactit atgetteegg etegtatgit 6180
      gtgtggaatt gtgagcggat aacaatttca cacaggaaac agctatgacc atgattacgc 6240
      caagetggeg eg
 15
      <210> 17
      <211> 25
      <212> DNA
      <213> Artificial Sequence
20
     <220>
      <223> Description of Artificial Sequence: primer P64-1
      <400> 17
     tcagcaacca ggctccccag caggc
                                                                           25
25
     <210> 18
     <211> 27
      <212> DNA
30
     <213> Artificial Sequence
     <223> Description of Artificial Sequence: primer 64-4
35
     <400> 18
     gacgacagta tcggcctcag gaagatc
                                                                          27
     <210> 19
40
     <211> 840
     <212> DNA
     <213> Artificial Sequence
     <220>
45
     <223> Description of Artificial Sequence:
           oligonucleotide 80d
     <400> 19
     ggtaccgagc tcggatcctc tagtaacggc cgccagtgtg ctggaattcg gcttcagcaa 60
50
     ccaggetece cageaggeag aagtatgeaa ageatgeate teaattagte ageaaceagg 120
     tgtggaaagt ccccaggete cccagcagge agaagtatge aaagcatgea teteaattag 180
     teageaacea tagteeegee cetaaeteeg eccateeege ecctaaetee geccagttee 240
     geccattete egececatgg etgactaatt ttttttattt atgeagagge egaggeegee 300
     teggeetagg aacagtegae gacactgeag agacetaett cactaacaae eggtacagtt 360
55
     cgtggaccag atgggtgagg tggagtacgc gcccggggag cccaaaggtt accccagttg 420
     gggcactact cccgaaaacc gcttctggat ccataacttc gtatagcata cattatacga 480
     agttataccg ggccaccatg gtcgcgagta gcttggcact ggggttgctt ttgcgnygtc 540
     gtgactggga aaaccctggc gttacccaac ttaatcgcct tgcagcacat ccccctttcg 600
     ccagctggcg taatagcgaa gaggcccgca ccgatcgccc ttcccaacag ttgcgcagct 660
60
     gaatggcgaa tggcgctttg cctggcttcc ggcaccagaa gcggtgccgg aaagctggct 720
     ggagtgcgat cttcctgagg ccgatactgt cgtcaagccg aattctgcag atatccatca 780
     cactggcggc cgctcgagca tgcatctaga gggccaattc gccctatagt gagtcgtatt 840
65
     <210> 20
     <211> 1842
     <212> DNA
```

<213> Bacteriophage phi-C31

<220> <221> CDS 5 <222> (1)..(1839)

<400> 20

35

atg aca caa ggg gtt gtg acc ggg gtg gac acg tac gcg ggt gct tac 48

Met Thr Gln Gly Val Val Thr Gly Val Asp Thr Tyr Ala Gly Ala Tyr

10 1 5 10 15

gac cgt cag tcg cgc gag cgc gag aat tcg agc gca gca agc cca gcg 96
Asp Arg Gln Ser Arg Glu Arg Glu Asn Ser Ser Ala Ala Ser Pro Ala
20 25 30

aca cag cgt agc gcc aac gaa gac aag gcg gcc gac ctt cag cgc gaa 144 Thr Gln Arg Ser Ala Asn Glu Asp Lys Ala Ala Asp Leu Gln Arg Glu 35

gtc gag cgc gac ggg ggc cgg ttc agg ttc gtc ggg cat ttc agc gaa 192
Val Glu Arg Asp Gly Gly Arg Phe Arg Phe Val Gly His Phe Ser Glu
50
55

gcg ccg ggc acg tcg gcg ttc ggg acg gcg gag cgc ccg gag ttc gaa 240
Ala Pro Gly Thr Ser Ala Phe Gly Thr Ala Glu Arg Pro Glu Phe Glu
65 70 75

cgc atc ctg aac gaa tgc cgc gcc ggg ccg ctc aac atg atc att gtc 288
Arg Ile Leu Asn Glu Cys Arg Ala Gly Arg Leu Asn Met Ile Ile Val
85 90 95

tat gac gtg tcg cgc ttc tcg cgc ctg aag gtc atg gac gcg att ccg 336 Tyr Asp Val Ser Arg Phe Ser Arg Leu Lys Val Met Asp Ala Ile Pro 100 105 110

att gtc tcg gaa ttg ctc gcc ctg ggc gtg acg att gtt tcc act cag 384
Ile Val Ser Glu Leu Leu Ala Leu Gly Val Thr Ile Val Ser Thr Gln
115 120 125

gaa ggc gtc ttc cgg cag gga aac gtc atg gac ctg att cac ctg att 432 Glu Gly Val Phe Arg Gln Gly Asn Val Met Asp Leu Ile His Leu Ile 130 135 140

atg cgg ctc gac gcg tcg cac aaa gaa tct tcg ctg aag tcg gcg aag 480

Met Arg Leu Asp Ala Ser His Lys Glu Ser Ser Leu Lys Ser Ala Lys
145

150

160

att ctc gac acg aag aac ctt cag cgc gaa ttg ggc ggg tac gtc ggc 528

Ile Leu Asp Thr Lys Asn Leu Gln Arg Glu Leu Gly Gly Tyr Val Gly

165 170 175

ggg aag gcg cct tac ggc ttc gag ctt gtt tcg gag acg aag gag atc 576 Gly Lys Ala Pro Tyr Gly Phe Glu Leu Val Ser Glu Thr Lys Glu Ile 180 185 190

acg cgc aac ggc cga atg gtc aat gtc gtc atc aac aag ctt gcg cac 624
Thr Arg Asn Gly Arg Met Val Asn Val Val Ile Asn Lys Leu Ala His
195 200 205

60 tcg acc act ccc ctt acc gga ccc ttc gag ttc gag ccc gac gta atc 672 Ser Thr Thr Pro Leu Thr Gly Pro Phe Glu Phe Glu Pro Asp Val Ile 210 215 220

cgg tgg tgg tgg cgt gag atc aag acg cac aaa cac ctt ccc ttc aag 720
Arg Trp Trp Arg Glu Ile Lys Thr His Lys His Leu Pro Phe Lys
225 230 235 240

WO 02/38613 ccg ggc agt caa gcc gcc att cac ccg ggc agc atc acg ggg ctt tgt 768 Pro Gly Ser Gln Ala Ala Ile His Pro Gly Ser Ile Thr Gly Leu Cys aag cgc atg gac gct gac gcc gtg ccg acc cgg ggc gag acg att ggg 816 Lys Arg Met Asp Ala Asp Ala Val Pro Thr Arg Gly Glu Thr Ile Gly 260 265 aag aag acc gct toa agc gcc tgg gac ccg gca acc gtt atg cga atc 864 10 Lys Lys Thr Ala Ser Ser Ala Trp Asp Pro Ala Thr Val Met Arg Ile ctt cgg gac ccg cgt att gcg ggc ttc gcc gct gag gtg atc tac aag 912 Leu Arg Asp Pro Arg Ile Ala Gly Phe Ala Ala Glu Val Ile Tyr Lys 15 aag aag ccg gac ggc acg ccg acc acg aag att gag ggt tac cgc att 960 Lys Lys Pro Asp Gly Thr Pro Thr Thr Lys Ile Glu Gly Tyr Arg Ile 310 20 cag cgc gac ccg atc acg ctc cgg ccg gtc gag ctt gat tgc gga ccg 1008 Gln Arg Asp Pro Ile Thr Leu Arg Pro Val Glu Leu Asp Cys Gly Pro 25 atc atc gag ccc gct gag tgg tat gag ctt cag gcg tgg ttg gac ggc 1056 Ile Ile Glu Pro Ala Glu Trp Tyr Glu Leu Gln Ala Trp Leu Asp Gly agg ggg cgc ggc aag ggg ctt tcc cgg ggg caa gcc att ctg tcc gcc 1104 30 Arg Gly Arg Gly Lys Gly Leu Ser Arg Gly Gln Ala Ile Leu Ser Ala 360 atg gac aag ctg tac tgc gag tgt ggc gcc gtc atg act tcg aag cgc 1152 -Met Asp Lys Leu Tyr Cys Glu Cys Gly Ala Val Met Thr Ser Lys Arg 35 370 ggg gaa gaa tog ato aag gao tot tac ogc tgc ogt ogc ogg aag gtg 1200 Gly Glu Glu Ser Ile Lys Asp Ser Tyr Arg Cys Arg Arg Arg Lys Val 390 40 gtc gac ccg tcc gca cct ggg cag cac gaa ggc acg tgc aac gtc agc Val Asp Pro Ser Ala Pro Gly Gln His Glu Gly Thr Cys Asn Val Ser 405 410 45 atg gcg gca ctc gac aag ttc gtt gcg gaa cgc atc ttc aac aag atc 1296 Met Ala Ala Leu Asp Lys Phe Val Ala Glu Arg Ile Phe Asn Lys Ile agg cac gec gaa gge gac gaa gag acg ttg geg ett etg tgg gaa gee 1344 50 Arg His Ala Glu Gly Asp Glu Glu Thr Leu Ala Leu Leu Trp Glu Ala 440 gee ega ege tte gge aag ete aet gag geg eet gag aag age gge gaa 1392 Ala Arg Arg Phe Gly Lys Leu Thr Glu Ala Pro Glu Lys Ser Gly Glu 55 cgg gcg aac ctt gtt gcg gag cgc gcc gac gcc ctg aac gcc ctt gaa 1440 Arg Ala Asn Leu Val Ala Glu Arg Ala Asp Ala Leu Asn Ala Leu Glu 60 gag ctg tac gaa gac cgc gcg gca ggc gcg tac gac gga ccc gtt ggc 1488 Glu Leu Tyr Glu Asp Arg Ala Ala Gly Ala Tyr Asp Gly Pro Val Gly 485

agg aag cac ttc cgg aag caa cag gca gcg ctg acg ctc cgg cag caa Arg Lys His Phe Arg Lys Gln Gln Ala Ala Leu Thr Leu Arg Gln Gln 505

| 5 | GJ A G B B B B B B B B B B B B B B B B B B B | gcg Ala | gaa Glu 515 | gag Glu | cgg Arg | ctt Leu | gcc Ala | gaa Glu 520 | ctt Leu | gaa Glu | gcc Ala | gcc Ala | gaa Glu 525 | gcc Ala | ccg Pro | aag Lys | 1584 |
|----------------|---|--|------------------------------------|--|---|--|--|--|--|--|---------------------------------------|--|--|---|---|--------------------------------|------|
| | ctt Leu | ccc Pro 530 | ctt Leu | gac Asp | caa Gln | tgg Trp | ttc Phe 535 | ccc Pro | gaa Glu | gac Asp | gcc Ala | gac Asp 540 | gct Ala | gac Asp | ccg Pro | acc Thr | 1632 |
| 10 | ggc Gly 545 | Pro | aag Lys | tcg Ser | tgg Trp | tgg Trp 550 | ggg Gly | cgc Arg | gcg Ala | tca Ser | gta Val 555 | gac Asp | gac Asp | aag Lys | cgc Arg | gtg Val 560 | 1680 |
| 15 | ttc Phe | gtc Val | Gly | ctc Leu | ttc Phe 565 | gta Val | gac Asp | aag Lys | atc Ile | gtt Val 570 | gtc Val | acg Thr | aag Lys | tcg Ser | act Thr 575 | acg Thr | 1728 |
| 20 | Gly | agg Arg | GJ À âàà | cag Gln 580 | gga Gly | acg Thr | ccc Pro | atc Ile | gag Glu 585 | aag Lys | cgc Arg | gct Ala | tcg Ser | atc Ile 590 | acg Thr | tgg Trp | 1776 |
| 25 · | gcg Ala | aag Lys | ccg Pro 595 | ccg Pro | acc Thr | gac Asp | gac Asp | gac Asp 600 | gaa Glu | gac Asp | gac Asp | gcc Ala | cag Gln 605 | gac Asp | ggc Gly | acg Thr | 1824 |
| | gaa Glu | gac Asp 610 | gta Val | gcg Ala | gcg Ala | tag | | | | | | | | | | | 1842 |
| 30 | <212 <212 | 0> 2: L> 6: 2> PI 3> Ba | 13 | iopl | nage | phi- | -C31 | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| 35 | | | | • | , | • | | | | | | | | | | | |
| | <400 Met 1 |)> 2: Thr | l Gln | Gly | Val 5 | Val | Thr | | | 10 | | | | | 15 | | |
| 35 40 | <400 Met 1 Asp |)> 2: Thr Arg | l Gln Gln | Gly Ser 20 | Val 5 Arg | Val Glu | Thr Arg | Glu | Asn 25 | 10 Ser | Ser | Ala | Ala | Ser 30 | 15 Pro | Ala | |
| | <400 Met 1 Asp |)> 2: Thr Arg Gln | Gln Gln Arg 35 | Gly Ser 20 Ser | Val 5 Arg | Val Glu Asn | Thr Arg Glu | Glu Asp 40 | Asn 25 Lys | 10 Ser Ala | Ser Ala | Ala Asp | Ala Leu 45 | Ser 30 Gln | 15 Pro Arg | Ala Glu | · |
| 40 | <400 Met 1 Asp |)> 2: Thr Arg Gln | Gln Gln Gln Arg | Gly Ser 20 Ser | Val 5 Arg | Val Glu Asn | Thr Arg Glu | Glu Asp 40 | Asn 25 Lys | 10 Ser Ala | Ser Ala | Ala Asp | Ala Leu 45 | Ser 30 Gln | 15 Pro Arg | Ala Glu | |
| 40 | <400 Met 1 Asp Thr Val | Thr Arg Gln Glu 50 Pro | Gln Gln Arg 35 Arg Gly | Gly Ser 20 Ser Asp | Val 5 Arg Ala Gly Ser | Val Glu Asn Gly Ala 70 | Thr Arg Glu Arg 55 Phe | Glu Asp 40 Phe Gly | Asn 25 Lys Arg | 10 Ser Ala Phe | Ser Ala Val Glu 75 | Ala Asp Gly 60 Arg | Ala Leu 45 His | Ser 30 Gln Phe | 15 Pro Arg Ser | Ala Glu Glu Glu 80 | |
| 40 45 | <400 Met 1 Asp Thr Val | Thr Arg Gln Glu 50 Pro | Gln Gln Arg 35 Arg | Gly Ser 20 Ser Asp | Val 5 Arg Ala Gly Ser | Val Glu Asn Gly Ala 70 | Thr Arg Glu Arg 55 Phe | Glu Asp 40 Phe Gly | Asn 25 Lys Arg | 10 Ser Ala Phe | Ser Ala Val Glu 75 | Ala Asp Gly 60 Arg | Ala Leu 45 His | Ser 30 Gln Phe | 15 Pro Arg Ser | Ala Glu Glu Glu 80 | |
| 40 45 | <400 Met 1 Asp Thr Val Ala 65 Arg | Thr Arg Gln Glu 50 Pro | Gln Gln Arg 35 Arg Gly | Gly Ser 20 Ser Asp Thr | Val 5 Arg Ala Gly Ser | Val Glu Asn Gly Ala 70 Cys | Thr Arg Glu Arg 55 Phe | Glu Asp 40 Phe Gly Ala | Asn 25 Lys Arg Thr | 10 Ser Ala Phe Ala Arg 90 | Ser Ala Val Glu 75 Leu | Ala Asp Gly 60 Arg | Ala Leu 45 His Pro | Ser 30 Gln Phe Glu | 15 Pro Arg Ser Phe | Ala Glu Glu Glu 80 Val | |
| 40 45 50 | <400 Met 1 Asp Thr Val Ala 65 Arg | Thr Arg Gln Glu 50 Pro Ile Asp | Gln Gln Arg 35 Arg Gly Leu | Gly Ser 20 Ser Asp Thr Asn Ser 100 | Val 5 Arg Ala Gly Ser Glu 85 | Val Glu Asn Gly Ala 70 Cys | Thr Arg Glu Arg 55 Phe Arg | Glu Asp 40 Phe Gly Ala Arg | Asn 25 Lys Arg Thr Gly Leu 105 | 10 Ser Ala Phe Ala Arg 90 Lys | Ser Ala Val Glu 75 Leu Val | Ala Asp Gly 60 Arg Asn Met | Ala Leu 45 His Pro Met | Ser 30 Gln Phe Glu Ile Ala 110 | 15 Pro Arg Ser Phe Ile 95 Ile | Ala Glu Glu 80 Val | |
| 40 45 50 | <400 Met 1 Asp Thr Val Ala 65 Arg Tyr | Thr Arg Gln Glu 50 Pro Ile Asp Val | Gln Gln Arg 35 Arg Gly Leu Val Ser | Gly Ser 20 Ser Asp Thr Asn Ser 100 Glu | Val 5 Arg Ala Gly Ser Glu 85 Arg | Val Glu Asn Gly Ala 70 Cys Phe | Thr Arg Glu Arg 55 Phe Arg Ser | Glu Asp 40 Phe Gly Ala Arg Leu 120 | Asn 25 Lys Arg Thr Gly Leu 105 Gly | 10 Ser Ala Phe Ala Arg 90 Lys | Ser Ala Val Glu 75 Leu Val | Ala Asp Gly 60 Arg Asn Met | Ala Leu 45 His Pro Met Asp Val 125 | Ser 30 Gln Phe Glu Ile Ala 110 Ser | 15 Pro Arg Ser Phe Ile 95 Ile | Ala Glu Glu 80 Val Pro Gln | |
| 40 45 50 | <400 Met 1 Asp Thr Val Ala 65 Arg Tyr Ile | Thr Arg Gln Glu 50 Pro Ile Asp Val Gly 130 | Gln Gln Arg 35 Arg Gly Leu Val Ser | Gly Ser 20 Ser Asp Thr Asn Ser 100 Glu Phe | Val 5 Arg Ala Gly Ser Glu 85 Arg Leu | Val Glu Asn Gly Ala 70 Cys Phe Leu Gln | Thr Arg Glu Arg 55 Phe Arg Ser Ala Gly 135 | Glu Asp 40 Phe Gly Ala Arg Leu 120 Asn | Asn 25 Lys Arg Thr Gly Leu 105 Gly | 10 Ser Ala Phe Ala Arg 90 Lys Val Met | Ser Ala Val Glu 75 Leu Val Thr | Ala Asp Gly 60 Arg Asn Met Ile Leu 140 | Ala Leu 45 His Pro Met Asp Val 125 Ile | Ser 30 Gln Phe Glu Ile Ala 110 Ser His | 15 Pro Arg Ser Phe Ile 95 Ile Thr | Ala Glu Glu 80 Val Pro Gln Ile | |

| | | | | _ | | | | | | | | | | | | 101/ |
|----|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | | | | | 165 | | | | | 18 170 | | | | | 175 | |
| 5 | Gly | Lys | Ala | Pro 180 | Tyr | Gly | Phe | Ģlu | Leu 185 | Val | Ser | Glu | Thr | Lys 190 | Glu | Ile |
| - | Thr | Arg | Asn 195 | Gly | Arg | Met | Val | Asn 200 | Val | Val | Ile | Asn | Lys 205 | Leu | Ala | His |
| 10 | Ser | Thr 210 | Thr | Pro | Leu | Thr | Gly 215 | Pro | Phe | Glu | Phe | Glu 220 | Pro | Asp | Val | Ile |
| | Arg 225 | Trp | Trp | Trp | Arg | Glu 230 | Ile | Lys | Thr | His | Lys 235 | His | Leu | Pro | Phe | Lys 240 |
| 15 | Pro | Gly | Ser | Gln | Ala 245 | Ala | Ile | His | Pro | Gly 250 | Ser | Ile | Thr | Gly | Leu 255 | Cys |
| 20 | Lys | Arg | Met | Asp 260 | Ala | Asp | Ala | Val | Pro 265 | Thr | Arg | Gly | Glu | Thr 270 | Ile | Gly |
| 20 | Lys | Lys | Thr 275 | Ala | Ser | Ser | Ala | Trp 280 | Asp | Pro | Ala | Thr | Val 285 | Met | Arg | Ile |
| 25 | Leu | Arg 290 | Asp | Pro | Arg | Ile | Ala 295 | Gly | Phe | Ala | Ala | Glu 300 | Val | Ile | Tyr | Lys |
| | Lys 305 | Lys | Pro | Asp | Gly | Thr 310 | Pro | Thr | Thr | Lys | Ile 315 | Glu | Gly | Tyr | Arg | Ile 320 |
| 30 | Gln | Arg | Asp | Pro | Ile 325 | Thr | Leu | Arg | Pro | Va1 330 | Glu | Leu | Asp | Cys | Gly 335 | Pro |
| 35 | Ile | Ile | Glu | Pro 340 | Ala | Glu | Trp | Tyr | Glu 345 | Leu | Gln | Ala | Trp | Leu 350 | Asp | Gly |
| | Arg | Gly | Arg 355 | Gly | Lys | Gly | Leu | Ser 360 | Arg | Gly | Gln | Ala | Ile 365 | Leu | Ser | Ala |
| 40 | Met | Asp 370 | Lys | Leu | Tyr | Cys | Glu 375 | Cys | Gly | Ala : | Val | Met 380 | Thr | Ser | Lys | Arg |
| | Gly 385 | Glu | Glu | Ser | Ile | Lys 390 | Asp | Ser | Tyr | Arg | Cys 395 | Arg | Arg | Arg | Lys | Val 400 |
| 45 | Val | Asp | Pro | Ser | Ala 405 | Pro | Gly | Gln | His | Glu 410 | Gly | Thr | Cys | Asn | Val 415 | Ser |
| 50 | Met | Ala | Ala | Leu 420 | Asp | Lys | Phe | Val | Ala 425 | Glu | Arg | Ile | Phe | Asn 430 | Lys | Ile |
| | Arg | His | Ala 435 | Glu | Gly | Asp | Glu | Glu 440 | Thr | Leu | Ala | Leu | Leu 445 | Trp | Glu | Ala |
| 55 | Ala | Arg 450 | Arg | Phe | Gly | Lys | Leu 455 | Thr | Glu | Ala | Pro | Glu 460 | Lys | Ser | Gly | Glu |
| | Arg 465 | Ala | Asn | Leu | Val | Ala 470 | Glu | Arg | Ala | Asp | Ala 475 | Leu | Asn | Ala | Leu | Glu 480 |
| 60 | Glu | Leu | Tyr | Glu | Asp 485 | Arg | Ala | Ala | Gly | Ala 490 | Tyr | Asp | Gly | Pro | Val 495 | Gly |
| 65 | Arg | Lys | His | Phe 500 | Arg | Lys | Gln | Gln | Ala 505 | Ala | Leu | Thr | Leu | Arg 510 | Gln | Gln |
| | Gly | Ala | Glu 515 | Glu | Arg | Leu | Ala | Glu 520 | Leu | Glu | Ala | Ala | Glu 525 | Ala | Pro | Lys |
| | | | | | | | | | | | | | | | | |

| | Leu | Pro 530 | Leu | Asp | Gln | Trp | Phe 535 | Pro | Glu | Asp | Ala | Asp 540 | Ala | Asp | Pro | Thr | |
|----|------------------|----------------------------------|-------------------|-------------------|------------------|------------------|------------------|-------------------|-------------------|------------------|------------------|-------------|-------------------|-------------------|------------------|------------------|-----|
| 5 | Gly 545 | Pro | Lys | Ser | Trp | Trp 550 | Gly | Arg | Ala | Ser | Val 555 | Asp | Asp | Lys | Arg | Val 560 | |
| 10 | Phe | Val | Gly | Leu | Phe 565 | Val | Asp | Lys | Ile | Val 570 | Val | Thr | Lys | Ser | Thr 575 | Thr | |
| •• | Gly | Arg | Gly | Gln 580 | Gly | Thr | Pro | Ile | Glu 585 | Lys | Arg | Ala | Ser | Ile 590 | Thr | Trp | |
| 15 | Ala | Lys | Pro 595 | Pro | Thr | Asp | Asp | Asp 600 | Glu | Asp | Asp | Ala | Gln 605 | Asp | Gly | Thr | |
| | Glu | Asp 610 | Val | Ala | Ala | | | | | | | | | | | | |
| 20 | | | | | | | | | | | | | | | | • | |
| 25 | <21: | 0> 2: 1> 1: 2> Di 3> A: | 863 NA | icial | l Sed | quen | ce | | | | | | | | | | |
| 30 | | 3> De | escri | iptio | on of | f Art | tific prot | cial tein | Sequ C31- | ience -Int | e: Di (CNL | NA se S) | | nce | | | |
| | | L> CI | | (1860 |)) | | | | | | | | | | | | |
| 35 | <400 |)> 22 | 2 | | | | | | | | | | | | | | |
| | atg Met 1 | aca Thr | caa Gln | Gly ggg | gtt Val 5 | gtg Val | acc Thr | Gly | gtg Val | gac Asp 10 | acg Thr | tac Tyr | gcg Ala | ggt Gly | gct Ala 15 | tac Tyr | 48 |
| 40 | gac Asp | cgt Arg | cag Gln | tcg Ser 20 | cgc Arg | gag Glu | cgc Arg | gag Glu | aat Asn 25 | tcg Ser | agc Ser | gca Ala | gca Ala | agc Ser 30 | cca Pro | gcg Ala | 96 |
| 45 | aca Thr | cag Gln | cgt Arg 35 | agc Ser | gcc Ala | aac Asn | gaa Glu | gac Asp 40 | aag Lys | gcg Ala | gcc Ala | gac Asp | ctt Leu 45 | cag Gln | cgc Arg | gaa Glu | 144 |
| 50 | gtc Val | gag Glu 50 | cgc Arg | gac Asp | GJÀ aaa | ggc Gly | cgg Arg 55 | ttc Phe | agg Arg | ttc Phe | gtc Val | 60 G1y | cat His | ttc Phe | agc Ser | gaa Glu | 192 |
| 55 | gcg Ala 65 | ccg Pro | ggc Gly | acg Thr | tcg Ser | gcg Ala 70 | ttc Phe | GJA āāā | acg Thr | gcg Ala | gag Glu 75 | cgc Arg | ccg Pro | gag Glu | ttc Phe | gaa Glu 80 | 240 |
| | cgc Arg | atc Ile | ctg Leu | aac Asn | gaa Glu 85 | tgc Cys | cgc Arg | gcc Ala | ggg Gly | cgg Arg 90 | ctc Leu | aac Asn | atg Met | atc Ile | att Ile 95 | gtc Val | 288 |
| 60 | tat Tyr | gac Asp | gtg Val | tcg Ser 100 | cgc Arg | ttc Phe | tcg Ser | cgc Arg | ctg Leu 105 | aag Lys | gtc Val | atg Met | gac Asp | gcg Ala 110 | att Ile | ccg Pro | 336 |
| 65 | att Ile | gtc Val | tcg Ser 115 | gaa Glu | ttg Leu | ctc Leu | gcc Ala | ctg Leu 120 | ggc Gly | gtg Val | acg Thr | att Ile | gtt Val 125 | tcc Ser | act Thr | cag Gln | 384 |

cgg tgg tgg tgg cgt gag atc aag acg cac aaa cac ctt ccc ttc aag 720 Arg Trp Trp Trp Arg Glu Ile Lys Thr His Lys His Leu Pro Phe Lys 230 eeg gge agt caa gee gee att cae eeg gge age ate aeg ggg ett tgt 768 30 Pro Gly Ser Gln Ala Ala Ile His Pro Gly Ser Ile Thr Gly Leu Cys 250 aag cgc atg gac gct gac gcc gtg ccg acc cgg ggc gag acg att ggg 816 Lys Arg Met Asp Ala Asp Ala Val Pro Thr Arg Gly Glu Thr Ile Gly 260 aag aag acc gct tca agc gcc tgg gac ccg gca acc gtt atg cga atc 864 Lys Lys Thr Ala Ser Ser Ala Trp Asp Pro Ala Thr Val Met Arg Ile 280 40 ctt cgg gac ccg cgt att gcg ggc ttc gcc gct gag gtg atc tac aag Leu Arg Asp Pro Arg Ile Ala Gly Phe Ala Ala Glu Val Ile Tyr Lys 300 45 aag aag eeg gae gge aeg eeg aee aeg aag att gag ggt tae ege att 960 Lys Lys Pro Asp Gly Thr Pro Thr Thr Lys Ile Glu Gly Tyr Arg Ile 310 cag cgc gac ccg atc acg ctc cgg ccg gtc gag ctt gat tgc gga ccg 1008 50 Gln Arg Asp Pro Ile Thr Leu Arg Pro Val Glu Leu Asp Cys Gly Pro 330 atc atc gag ccc get gag tgg tat gag ett cag geg tgg ttg gac gge 1056 Ile Ile Glu Pro Ala Glu Trp Tyr Glu Leu Gln Ala Trp Leu Asp Gly - 55 agg ggg cgc ggc aag ggg ctt tee egg ggg caa gee att etg tee gee 1104 Arg Gly Arg Gly Lys Gly Leu Ser Arg Gly Gln Ala Ile Leu Ser Ala 355 60 atg gac aag ctg tac tgc gag tgt ggc gcc gtc atg act tcg aag cgc 1152 Met Asp Lys Leu Tyr Cys Glu Cys Gly Ala Val Met Thr Ser Lys Arg 370 65 ggg gaa gaa tog ato aag gao tot tao ogo tgo ogt ogo ogg aag gtg Gly Glu Glu Ser Ile Lys Asp Ser Tyr Arg Cys Arg Arg Arg Lys Val 395

| 5 | gtc ga Val As | c ccg p Pro | tcc Ser | gca Ala 405 | Pro | GJ A ā ā ā | cag Gln | cac His | gaa Glu 410 | Gly | acg Thr | tgc Cys | aac Asn | gtc Val 415 | agc Ser | 1248 |
|----|---|----------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|------|
| | atg gc Met Al | g gca a Ala | ctc Leu 420 | ASP | aag Lys | ttc Phe | gtt Val | gcg Ala 425 | gaa Glu | cgc Arg | atc Ile | ttc Phe | aac Asn 430 | aag Lys | atc Ile | 1296 |
| 10 | agg ca Arg Hi | c gcc s Ala 435 | GIU | ggc Gly | gac Asp | gaa Glu | gag Glu 440 | acg Thr | ttg Leu | gcg Ala | ctt Leu | ctg Leu 445 | tgg Trp | gaa Glu | gcc Ala | 1344 |
| 15 | gcc cg Ala Ar 45 | y Arg | ttc Phe | ggc Gly | aag Lys | ctc Leu 455 | act Thr | gag Glu | gcg Ala | cct Pro | gag Glu 460 | aag Lys | agc Ser | ggc Gly | gaa Glu | 1392 |
| 20 | cgg gc Arg Al 465 | g aac a Asn | ctt Leu | gtt Val | gcg Ala 470 | gag Glu | cgc Arg | gcc Ala | gac Asp | gcc Ala 475 | ctg Leu | aac Asn | gcc Ala | ctt Leu | gaa Glu 480 | 1440 |
| 25 | gag cte Glu Le | g tac ı Tyr | gaa Glu | gac Asp 485 | cgc Arg | gcg Ala | gca Ala | Gly | gcg Ala 490 | tac Tyr | gac Asp | gga Gly | ccc Pro | gtt Val 495 | ggc Gly | 1488 |
| | agg aag Arg Lys | g cac s His | ttc Phe 500 | cgg Arg | aag Lys | caa Gln | cag Gln | gca Ala 505 | gcg Ala | ctg Leu | acg Thr | ctc Leu | cgg Arg 510 | cag Gln | caa Gln | 1536 |
| 30 | Gly Ala | gaa Glu 515 | gag Glu | cgg Arg | ctt Leu | gcc Ala | gaa Glu 520 | ctt Leu | gaa Glu | gcc Ala | gcc Ala | gaa Glu 525 | gcc Ala | ccg Pro | aag Lys | 1584 |
| 35 | Ctt CCC Leu Pro 530 |) ren | gac Asp | caa Gln | tgg Trp | ttc Phe 535 | ccc Pro | gaa Glu | gac Asp | gcc Ala | gac Asp 540 | gct Ala | gac Asp | ccg Pro | acc Thr | 1632 |
| 40 | ggc cct Gly Pro 545 | aag Lys | tcg Ser | tgg Trp | tgg Trp 550 | ggg Gly | cgc Arg | gcg Ala | tca Ser | gta Val 555 | gac Asp | gac Asp | aag Lys | cgc Arg | gtg Val 560 | 1680 |
| 45 | ttc gtc Phe Val | Gly ggg | ctc Leu | ttc Phe 565 | gta Val | gac Asp | aag Lys | atc Ile | gtt Val 570 | gtc Val | acg Thr | aag Lys | tcg Ser | act Thr 575 | acg Thr | 1728 |
| | ggc agg Gly Arg | Gly | cag Gln 580 | gga Gly | acg Thr | ccc Pro | atc Ile | gag Glu 585 | aag Lys | cgc Arg | gct Ala | tcg Ser | atc Ile 590 | acg Thr | tgg Trp | 1776 |
| 50 | gcg aag Ala Lys | ccg Pro 595 | ccg Pro | acc Thr | gac Asp | gac Asp | gac Asp 600 | gaa Glu | gac Asp | gac Asp | gcc Ala | cag Gln 605 | gac Asp | ggc Gly | acg Thr | 1824 |
| 55 | gaa gac Glu Asp 610 | val | gcg Ala | gcg Ala | cct Pro | aag Lys 615 | aag Lys | aag Lys | agg Arg | aag Lys | gtt Val 620 | tag | | | | 1863 |
| 60 | <210> 2 <211> 6 <212> P <213> A <223> D | 20 RT rtifi escri | .ptio | n of | Art | ific | ial | Sequ | ence | : DN | A se | auen | Ce | | | |
| 65 | <400> 2 Met Thr | oaing | IOI | rus | lon | prot | ein | C31- | Int(| CNLS |) | | | Ala | Tyr | |

22 10 Asp Arg Gln Ser Arg Glu Arg Glu Asn Ser Ser Ala Ala Ser Pro Ala 5 Thr Gln Arg Ser Ala Asn Glu Asp Lys Ala Ala Asp Leu Gln Arg Glu Val Glu Arg Asp Gly Gly Arg Phe Arg Phe Val Gly His Phe Ser Glu 10 Ala Pro Gly Thr Ser Ala Phe Gly Thr Ala Glu Arg Pro Glu Phe Glu Arg Ile Leu Asn Glu Cys Arg Ala Gly Arg Leu Asn Met Ile Ile Val Tyr Asp Val Ser Arg Phe Ser Arg Leu Lys Val Met Asp Ala Ile Pro 20 Ile Val Ser Glu Leu Leu Ala Leu Gly Val Thr Ile Val Ser Thr Gln Glu Gly Val Phe Arg Gln Gly Asn Val Met Asp Leu Ile His Leu Ile Met Arg Leu Asp Ala Ser His Lys Glu Ser Ser Leu Lys Ser Ala Lys 30 Ile Leu Asp Thr Lys Asn Leu Gln Arg Glu Leu Gly Gly Tyr Val Gly 170 Gly Lys Ala Pro Tyr Gly Phe Glu Leu Val Ser Glu Thr Lys Glu Ile 35 Thr Arg Asn Gly Arg Met Val Asn Val Val Ile Asn Lys Leu Ala His Ser Thr Thr Pro Leu Thr Gly Pro Phe Glu Phe Glu Pro Asp Val Ile 40 Arg Trp Trp Trp Arg Glu Ile Lys Thr His Lys His Leu Pro Phe Lys 45 Pro Gly Ser Gln Ala Ala Ile His Pro Gly Ser Ile Thr Gly Leu Cys Lys Arg Met Asp Ala Asp Ala Val Pro Thr Arg Gly Glu Thr Ile Gly 50 Lys Lys Thr Ala Ser Ser Ala Trp Asp Pro Ala Thr Val Met Arg Ile Leu Arg Asp Pro Arg Ile Ala Gly Phe Ala Ala Glu Val Ile Tyr Lys 55 Lys Lys Pro Asp Gly Thr Pro Thr Thr Lys Ile Glu Gly Tyr Arg Ile 60 Gln Arg Asp Pro Ile Thr Leu Arg Pro Val Glu Leu Asp Cys Gly Pro Ile Ile Glu Pro Ala Glu Trp Tyr Glu Leu Gln Ala Trp Leu Asp Gly 65

Arg Gly Arg Gly Lys Gly Leu Ser Arg Gly Gln Ala Ile Leu Ser Ala

| | Met | Asp 370 | Lys. | Leu | Tyr | Суѕ | Glu 375 | Суз | Gly | Ala | Val | Met 380 | Thr | Ser | Lys | Arg |
|----|--------------|----------------------------------|-----------------|------------|------------|------------|------------|------------|-----------------|------------|------------|------------|------------|------------|------------|------------|
| 5 | Gly 385 | Glu | Glu | Ser | Ile | Lys 390 | Asp | Ser | Tyr | Arg | Cys 395 | Arg | Arg | Arg | Lys | Val 400 |
| 10 | Val | Asp | Pro | Ser | Ala 405 | Pro | Gly | Gln | His | Glu 410 | Gly | Thr | Cys | Asn | Val 415 | Ser |
| | Met | Ala | Ala | Leu 420 | Asp | Lys | Phe | Val | Ala 425 | Glu | Arg | Ile | Phe | Asn 430 | Lys | Ile |
| 15 | Arg | His | Ala 435 | Glu | Gly | Asp | Glu | Glu 440 | Thr | Leu | Ala | Leu | Leu 445 | Trp | Glu | Ala |
| | Ala | Arg 450 | Arg | Phe | Glà | Lys | Leu 455 | Thr | Glu | Ala | Pro | Glu 460 | Lys | Ser | Gly | Glu |
| 20 | Arg 465 | Ala | Asn | Leu | Val | Ala 470 | Glu | Arg | Ala | Asp | Ala 475 | Leu | Asn | Ala | Leu | Glu 480 |
| 25 | Glu | Leu | Tyr | Glu | Asp 485 | Arg | Ala | Ala | Gly | Ala 490 | Tyr | Asp | Gly | Pro | Val 495 | Gly |
| | Arg | Lys | His | Phe 500 | Arg | Lys | Gln | Gln | Ala 505 | Ala | Leu | Thr | Leu | Arg 510 | Gln | Gln |
| 30 | Gly | Ala | Glu 515 | Glu | Arg | Leu | Ala | Glu 520 | Leu | Glu | Ala | Ala | Glu 525 | Ala | Pro | Lys |
| | Leu | Pro 530 | Leu | Asp | Gln | Trp | Phe 535 | Pro | Glu | Asp | Ala | Asp 540 | Ala | Asp | Pro | Thr |
| 35 | Gly 545 | Pro | Lys | Ser | Trp | Trp 550 | Gly | Arg | Ala | Ser | Val 555 | Asp | Asp | Lys | Arg | Val 560 |
| 40 | Phe | Val | Gly | Leu | Phe 565 | Val | Asp | Lys | Ile | Val 570 | Val | Thr | ГÀЗ | Ser | Thr 575 | Thr |
| | Gly | Arg | Gly | Gln 580 | Gly | Thr | Pro | Ile | Glu 585 | Lys | Arg | Ala | Ser | Ile 590 | Thr | Trp |
| 45 | Ala | Lys | Pro 595 | Pro | Thr | Asp | Asp | Asp 600 | Glu | Asp | Asp | Ala | Gln 605 | Asp | Gly | Thr |
| | Glu | Asp 610 | Val | Ala | Ala | Pro | Lys 615 | Lys | Lys | Arg | Lys | Val 620 | | | | |
| 50 | | | | | | | | | • | | | | | | | |
| 55 | <211 <212 |)> 24 l> 43 2> PF 3> Ar | 3 ? T | icial | L Sec | quenc | ce | | | | | | | | | |
| | <220 <223 | | escri | iptic | on of | f Art | ific | cial | Sequ | ience | e: NI | LS | | | | |
| 60 | |)> 24 Lys | | Lys | Lys 5 | Lys | Lys | Lуз | Lys | Lys 10 | Lys | Lys | Cys | Arg | Leu 15 | Lys |
| 65 | Lys | Leu | Lys | Cys 20 | Ser | Lys | Glu | Lys | Pro 25 | Lys | Cys | Ala | Lys | Cys 30 | Leu | Lys |
| | Lvs | Lvo | T.vo | T.179 | Δτα | Ara | Δτα | T.120 | ω) - | T | N | | | | | |

24

```
<210> 25
     <211> 10
     <212> PRT
     <213> Artificial Sequence
     <220>
10
   <223> Description of Artificial Sequence: NLS
     <400> 25
     Ile Lys Tyr Phe Lys Lys Phe Pro Lys Asp
15
     <210> 26
     <211> 14
     <212> PRT
20
   <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: NLS
25
     <400> 26
     Met Thr Gly Ser Lys Thr Arg Lys His Arg Gly Ser Gly Ala
30
     <210> 27
     <211> 14
     <212> PRT
     <213> Artificial Sequence
35
     <220>
     <223> Description of Artificial Sequence: NLS
     <400> 27
     Met Thr Gly Ser Lys His Arg Lys His Pro Gly Ser Gly Ala
40
                       5
     <210> 28
     <211> 7
45
     <212> PRT
     <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: NLS
50
     <400> 28
     Gly Lys Lys Arg Ser Lys Ala
55
     <210> 29
     <211> 14
     <212> PRT
     <213> Artificial Sequence
60
     <220>
     <223> Description of Artificial Sequence: NLS
     <400> 29
65
     Pro Lys Lys Ala Arg Glu Asp Val Ser Arg Lys Arg Pro Arg
```

```
<210> .30
     <211> 11
     <212> PRT
    <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: NLS
10 <400> 30
     Ala Pro Lys Arg Lys Ser Gly Val Ser Lys Cys
15
     <210> 31
     <211> 12
     <212> PRT
     <213> Artificial Sequence
20
     <220>
     <223> Description of Artificial Sequence: NLS
     <400> 31
     Glu Glu Asp Gly Pro Gln Lys Lys Lys Arg Arg Leu
25
     <210> 32
     <211> 8
     <212> PRT
     <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: NLS
35
     <400> 32
     Ala Pro Thr Lys Arg Lys Gly Ser
40
     <210> 33
     <211> 7
     <212> PRT
     <213> Artificial Sequence
45
     <220>
     <223> Description of Artificial Sequence: NLS
     <400> 33
50
     Pro Asn Lys Lys Arg Lys
     <210> 34
55
     <211> 5
     <212> PRT
     <213> Artificial Sequence
     <220>
60
     <223> Description of Artificial Sequence: NLS
     <400> 34
     Lys Arg Pro Arg Pro
65
     <210> 35
```

<211> 11 <212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: NLS

<400> 35

Cys Gly Gly Leu Ser Ser Lys Arg Pro Arg Pro 10

<210> 36

<211> 19

<212> PRT

<213> Artificial Sequence

<223> Description of Artificial Sequence: NLS 20

<400> 36

Pro Pro Lys Lys Arg Met Arg Arg Ile Glu Pro Lys Lys Lys

25 Lys Arg Pro

<210> 37 30

<211> 11

<212> PRT <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: NLS

<400> 37

Pro Phe Leu Asp Arg Leu Arg Arg Asp Gln Lys

40

<210> 38

<211> 9

<212> PRT

45 <213> Artificial Sequence

<223> Description of Artificial Sequence: NLS

50 <400> 38

Pro Lys Gln Lys Arg Lys Met Ala Arg

55 <210> 39

<211> 9

<212> PRT

<213> Artificial Sequence

60 <220>

<223> Description of Artificial Sequence: NLS

<400> 39

Ser Val Thr Lys Lys Arg Lys Leu Glu 65

<210> 40 <211> 11

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: NLS

<400> 40

10 Cys Gly Gly Ala Ala Lys Arg Val Lys Leu Asp 5

<210> 41

15 <211> 9 <212> PRT

<213> Artificial Sequence

<220>

20 <223> Description of Artificial Sequence: NLS

<400> 41

Pro Ala Ala Lys Arg Val Lys Leu Asp

25

<210> 42

<211> 11

<212> PRT

30 <213> Artificial Sequence

<223> Description of Artificial Sequence: NLS

<400> 42

Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro 5

40 <210> 43

<211> 8

<212> PRT

<213> Artificial Sequence

45 <220>

<223> Description of Artificial Sequence: NLS

<400> 43

Pro Gln Ser Arg Lys Lys Leu Arg 50

<210> 44

<211> 8

<212> PRT <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: NLS 60

<400> 44

Pro Leu Leu Lys Lys Ile Lys Gln

65

<210> 45

<211> 7

```
28
     <212> PRT
     <213> Artificial Sequence
    <223> Description of Artificial Sequence: NLS
     <400> 45
     Pro Gln Pro Lys Lys Pro
10
     <210> 46
     <211> 9
     <212> PRT
    <213> Artificial Sequence
     <223> Description of Artificial Sequence: NLS
20
    <400> 46 .
     Ser Lys Arg Val Ala Lys Arg Lys Leu
25
     <210> 47
    <211> 9
     <212> PRT
     <213> Artificial Sequence
30 <220>
    <223> Description of Artificial Sequence: NLS
    <400> 47
    Ala Ser Lys Ser Arg Lys Arg Lys Leu
35
     <210> 48
     <211> 16
    <212> PRT
    <213> Artificial Sequence
    <220>
    <223> Description of Artificial Sequence: NLS
45
     <400> 48
    Gly Gly Leu Cys Ser Ala Arg Leu His Arg His Ala Leu Leu Ala Thr
50
    <210> 49
    <211> 8
    <212> PRT
    <213> Artificial Sequence
55
    <220>
    <223> Description of Artificial Sequence: NLS
    <400> 49
60
    Arg Lys Thr Lys Lys Lys Ile Lys
                    5
    <210> 50
65
    <211> 8
    <212> PRT
```

<213> Artificial Sequence

```
<220>
      <223> Description of Artificial Sequence: NLS
      <400> 50
      Arg Lys Leu Lys Lys Leu Gly Asn
 10
      <210> 51
      <211> 8
      <212> PRT
      <213> Artificial Sequence
 15
      <220>
      <223> Description of Artificial Sequence: NLS
      <400> 51
      Arg Lys Asp Arg Arg Gly Gly Arg
 20
      <210> 52
      <211> 18
      <212> PRT
      <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: NLS
 30
      <400> 52
      Asp Thr Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg Leu Leu Arg Leu
 35
      Asp Glu
      <210> 53
 40
      <211> 7
      <212> PRT
      <213> Artificial Sequence
      <220> '
     <223> Description of Artificial Sequence: NLS
      Pro Lys Lys Arg Lys Val
 50
      <210> 54
      <211> 1410
      <212> DNA
55
     <213> Bacteriophage R4
      <220>
      <221> CDS
      <222> (1)..(1407)
60
      <400> 54
     atg aat cga ggg ggg ccc act gta cgg gcc gac atc tac gtc cga atc
     Met Asn Arg Gly Gly Pro Thr Val Arg Ala Asp Ile Tyr Val Arg Ile
65
     age etg gae ege aca ggg gaa gag ete ggg gte gag ege eag gag gag
                                                                         96
     Ser Leu Asp Arg Thr Gly Glu Glu Leu Gly Val Glu Arg Gln Glu Glu
```

| | | | | 20 | | | | | 25 | 30 | | | | 30 | | | |
|----|------------|------------------|------------|-------------------|------------|------------|------------------|------------|-------------------|------------|------------|------------------|------------|-------------------|------------|------------|-------------------|
| 5 | | | | gag Glu | | | | | | | | | | | | | 144 |
| 10 | tgg Trp | gtc Val 50 | gac Asp | aac Asn | gac Asp | ctg Leu | agc Ser 55 | gcc Ala | acc Thr | aag Lys | aag Lys | aac Asn 60 | gtc Val | gtc Val | cgc Arg | cct Pro | 192 |
| 10 | | | | gcg Ala | | | | | | | | | | | | | 240 |
| 15 | | | | cgg Arg | | | | | | | | | | | | | 288 |
| 20 | gac Asp | ctc Leu | gga Gly | gtc Val 100 | aac Asn | gtc Val | cac His | gcc Ala | gtg Val 105 | atg Met | gcc Ala | gga Gly | cac His | ctg Leu 110 | gac Asp | ctg Leu | 336 |
| 25 | | | | gcc Ala | | | | | | | | | | | | | 384 |
| 30 | | | | Gly ggc | | | | | | | | | | | | | 432 |
| | | | | cgc Arg | | | | | | | | | | | | | 480 |
| 35 | | | | gac Asp | | | | | | | | | | | | | 528 |
| 40 | | | | ggc Gly 180 | | | | | | | | | | | | | 576 |
| 45 | | | | tac Tyr | | | | | | | | | | | | | 624 |
| 50 | | | | ggc Gly | | | | | | | | | | | | | 672 |
| | | | | cgc Arg | | | | | | | | | | | | | 720 |
| 55 | | | | cag Gln | | | | | | | | | | | | | 7 [.] 68 |
| 60 | | | | atc Ile 260 | | | | | | | | | | | Pro | cgg Arg | 816 |
| 65 | | | | acc Thr | | | | | Leu | | | | | | | ggt Gly | 864 |
| | gag | tgc | ggc | aag | acg | gtc | agt | gga | cgc | ggc | tac | cga | ggt | gto | ctg | gtc | 912 |

| | • | | | _ | | | | | | | | | | | | 101/ | E1 01/12// |
|-------------|-------------------|----------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|------------|
| | Glu | Cys 290 | Gly | Lys | Thr | Val | Ser 295 | Gly | Arg | 31 Gly | Tyr | Arg 300 | Gly | Val | Leu | Val | |
| 5 | tac Tyr 305 | gga Gly | tgt Cys | aag Lys | gac Asp | acg Thr 310 | cac His | act Thr | cgg Arg | acg Thr | cct Pro 315 | cgg Arg | agc Ser | atc Ile | gct Ala | gac Asp 320 | 960 |
| 10 | ggc Gly | cgc Arg | gcg Ala | agc Ser | agc Ser 325 | tcg Ser | acc Thr | ctc Leu | gcc Ala | cgg Arg 330 | ctc Leu | atg Met | ttc Phe | ccc Pro | gac Asp 335 | ttc Phe | 1008 |
| 15 | ctg Leu | ccc Pro | ggc Gly | ctc Leu 340 | ctg Leu | gcc Ala | tct Ser | GJÅ āāā | cag Gln 345 | gcc Ala | gag Glu | gac Asp | ggc | cag Gln 350 | tcg Ser | gca Ala | 1056 |
| | gca Ala | tcc Ser | aag Lys 355 | cac His | tcg Ser | gag Glu | Ala | cag Gln 360 | acg Thr | ctg Leu | cgc Arg | gag Glu | cgc Arg 365 | ctt Leu | gac Asp | GJ y ggg | 1104 |
| 20 | ctg Leu | gct Ala 370 | acg Thr | gcc Ala | tac Tyr | gcg Ala | gag Glu 375 | ggt Gly | gcg Ala | atc Ile | agc Ser | ctg Leu 380 | tct Ser | cag Gln | atg Met | acg Thr | 1152 |
| 25 | gcc Ala 385 | ggc Gly | tcg Ser | gaa Glu | gca Ala | ctg Leu 390 | cgg Arg | aag Lys | aag Lys | ctg Leu | gag Glu 395 | gtg Val | atc Ile | gaa Glu | gcc Ala | gac Asp 400 | 1200 |
| 30 | ctc Leu | gtg Val | ggc Gly | tcg Ser | gca Ala 405 | ggc Gly | atc Ile | ccg Pro | ccc Pro | ttc Phe 410 | gat Asp | cca Pro | gtg Vaļ | gcc Ala | gga Gly 415 | gtg Val | 1248 |
| 35 . | gct Ala | ggc Gly | ctg Leu | atc Ile 420 | tcc Ser | ggc Gly | tgg Trp | ccc Pro | acc Thr 425 | acg Thr | cct Pro | ctc Leu | ccg Pro | acg Thr 430 | cgt Arg | cga Arg | 1296 |
| <i>33</i> . | gca Ala | tgg Trp | gtg Val 435 | gac Asp | ttc Phe | tgc Cys | ctg Leu | gtg Val 440 | gtc Val | acg Thr | ctg Leu | aac Asn | acc Thr 445 | cag Gln | aag Lys | Gly ggg | 1344 |
| 40 | cgc Arg | cat His 450 | gcg Ala | tcg Ser | agc Ser | atg Met | acc Thr 455 | gtg Val | gac Asp | gac Asp | cac His | gtc Val 460 | acc Thr | atc Ile | gag Glu | tgg Trp | 1392 |
| 45 | | | gtg Val | | | tag | | | | | | | | | | | 1410 |
| 50 | <21: | 0> 5! 1> 4! 2> P! 3> Ba | 69 | riopl | hage | R4 | | | | | | | • | | | | |
| 55 | | 0> 5! Asn | 5 Arg | Gly | Gly 5 | Pro | Thr | Val | Arg | Ala 10 | Asp | Ile | Tyr | Val | Arg 15 | | |
| 60 | Ser | Leu | Asp | Arg 20 | Thr | Gly | Glu | Glu | Leu 25 | Gly | Val | Glu | Arg | Gln 30 | | Glu | |
| | | | Arg 35 | | | | | 40 | • | | | | 45 | _ | | | |
| 65 | | 50 | | | | | 55 | | | | | 60 | | | | Pro | |
| | vsh | E116 | GTA | нта | Met | ттб | wrg | ser | ASTI | PIO | GIN | ΑΙΑ | ıте | val | cys | Trp | |

WO 02/38613 32 70 75 His Thr Asp Arg Leu Ile Arg Val Thr Arg Asp Leu Glu Arg Val Ile 5 Asp Leu Gly Val Asn Val His Ala Val Met Ala Gly His Leu Asp Leu Ser Thr Pro Ala Gly Arg Ala Val Ala Arg Thr Val Thr Ala Trp Ala 10 Thr Tyr Glu Gly Glu Gln Lys Ala Glu Arg Gln Lys Leu Ala Asn Ile 130 15 Gln Asn Ala Arg Ala Gly Lys Pro Tyr Thr Pro Gly Ile Arg Pro Phe Gly Tyr Gly Asp Asp His Met Thr Ile Val Thr Ala Glu Ala Asp Ala 20 Ile Arg Asp Gly Ala Lys Met Ile Leu Asp Gly Trp Ser Leu Ser Ala Val Ala Arg Tyr Trp Glu Glu Leu Lys Leu Gln Ser Pro Arg Ser Met 25 Ala Ala Gly Gly Lys Gly Trp Ser Leu Arg Gly Val Lys Lys Val Leu 210 215 220 30 Thr Ser Pro Arg Tyr Val Gly Arg Ser Ser Tyr Leu Gly Glu Val Val Gly Asp Ala Gln Trp Pro Pro Ile Leu Asp Pro Asp Val Tyr Tyr Gly 35 Val Val Ala Ile Leu Asn Asn Pro Asp Arg Phe Ser Gly Gly Pro Arg 265 Thr Gly Arg Thr Pro Gly Thr Leu Leu Ala Gly Ile Ala Leu Cys Gly 280 Glu Cys Gly Lys Thr Val Ser Gly Arg Gly Tyr Arg Gly Val Leu Val 295 -45 Tyr Gly Cys Lys Asp Thr His Thr Arg Thr Pro Arg Ser Ile Ala Asp Gly Arg Ala Ser Ser Ser Thr Leu Ala Arg Leu Met Phe Pro Asp Phe 330 50 Leu Pro Gly Leu Leu Ala Ser Gly Gln Ala Glu Asp Gly Gln Ser Ala Ala Ser Lys His Ser Glu Ala Gln Thr Leu Arg Glu Arg Leu Asp Gly 55 360 Leu Ala Thr Ala Tyr Ala Glu Gly Ala Ile Ser Leu Ser Gln Met Thr 60 Ala Gly Ser Glu Ala Leu Arg Lys Lys Leu Glu Val Ile Glu Ala Asp 395 Leu Val Gly Ser Ala Gly Ile Pro Pro Phe Asp Pro Val Ala Gly Val

410

Ala Gly Leu Ile Ser Gly Trp Pro Thr Thr Pro Leu Pro Thr Arg Arg

Ala Trp Val Asp Phe Cys Leu Val Val Thr Leu Asn Thr Gln Lys Gly 440 Arg His Ala Ser Ser Met Thr Val Asp Asp His Val Thr Ile Glu Trp 455 Arg Asp Val Ala Glu 10 <210> 56 <211> 1503 15 <212> DNA <213> CisA recombinase <220> <221> CDS 20 <222> (1)..(1500) <400> 56 gtg ata gca ata tat gta agg gta tcg acc gag gaa caa gcg atc aag Val Ile Ala Ile Tyr Val Arg Val Ser Thr Glu Glu Gln Ala Ile Lys 25 gga tcg agc atc gac agc caa atc gag gcc tgt ata aag aaa gca ggg 96 Gly Ser Ser Ile Asp Ser Gln Ile Glu Ala Cys Ile Lys Lys Ala Gly 25 30 act aaa gat gtg ctg aag tat gca gat gaa gga ttt tca gga gag ctt 144 Thr Lys Asp Val Leu Lys Tyr Ala Asp Glu Gly Phe Ser Gly Glu Leu 35 tta gaa cgt ccg gct ttg aat cgc ttg agg gag gat gca agc aag gga 192 Leu Glu Arg Pro Ala Leu Asn Arg Leu Arg Glu Asp Ala Ser Lys Gly ctt ata agt caa gtc att tgt tac gat cct gac cgt ctt tct cgg aaa 240 Leu Ile Ser Gln Val Ile Cys Tyr Asp Pro Asp Arg Leu Ser Arg Lys tta atg aat cag cta atc att gat gac gaa ttg cga aag cga aac ata 288 Leu Met Asn Gln Leu Ile Ile Asp Asp Glu Leu Arg Lys Arg Asn Ile 45 cct ttg att ttt gta aat ggt gaa tac gcc aat tct cca gaa ggt caa 336 Pro Leu Ile Phe Val Asn Gly Glu Tyr Ala Asn Ser Pro Glu Gly Gln 105 50 ttg ttt ttc gca atg cgc ggg gca atc tca gaa ttt gaa aaa gcc aaa Leu Phe Phe Ala Met Arg Gly Ala Ile Ser Glu Phe Glu Lys Ala Lys 55 atc aaa gaa cgg aca tca agc ggc cga ctt caa aaa atg aaa aaa ggc 432 Ile Lys Glu Arg Thr Ser Ser Gly Arg Leu Gln Lys Met Lys Lys Gly 130 135 atg atc att aaa gat tct aaa cta tat ggc tat aaa ttt gtt aaa gag 480 60 Met Ile Ile Lys Asp Ser Lys Leu Tyr Gly Tyr Lys Phe Val Lys Glu 145 150 aaa aga act ctt gag ata tta gaa gag gaa gca aaa atc att cgg atg Lys Arg Thr Leu Glu Ile Leu Glu Glu Glu Ala Lys Ile Ile Arg Met 65 att ttt aac tat ttc acc gat cat aaa agc cct ttt ttc ggc aga gta

34 Ile Phe Asn Tyr Phe Thr Asp His Lys Ser Pro Phe Phe Gly Arg Val 185 aat ggt att gct cta cat tta act cag atg ggg gtt aaa aca aaa aaa 624 Asn Gly Ile Ala Leu His Leu Thr Gln Met Gly Val Lys Thr Lys Lys ggc gcc aaa gta tgg cac agg cag gtt gtt cgg caa ata tta atg aac Gly Ala Lys Val Trp His Arg Gln Val Val Arg Gln Ile Leu Met Asn 10 215 tet tee tat aag ggt gaa cat aga cag tat aaa tat gat aca gag ggt 720 Ser Ser Tyr Lys Gly Glu His Arg Gln Tyr Lys Tyr Asp Thr Glu Gly tcc tat gtt tca aag cag gca ggg aac aaa tct ata att aaa ata agg 768 Ser Tyr Val Ser Lys Gln Ala Gly Asn Lys Ser Ile Ile Lys Ile Arg 20 cet gaa gaa gaa caa atc act gtg aca att cca gca att gtt cca gct 816 Pro Glu Glu Glu Gln Ile Thr Val Thr Ile Pro Ala Ile Val Pro Ala 265 gaa caa tgg gat tat gct caa gaa ctc tta ggt caa agt aaa aga aaa 864 Glu Gln Trp Asp Tyr Ala Gln Glu Leu Leu Gly Gln Ser Lys Arg Lys 275 cac ttg agt atc age cct cac aat tac ttg tta tcg ggt ttg gtt aga 912 His Leu Ser Ile Ser Pro His Asn Tyr Leu Leu Ser Gly Leu Val Arg 30 295 tgc gga aaa tgc gga aat acc atg aca ggg aag aaa aga aaa tca cat 960 Cys Gly Lys Cys Gly Asn Thr Met Thr Gly Lys Lys Arg Lys Ser His 310 35 ggt aaa gac tac tat gta tat act tgc cgg aaa aat tat tct ggc gca 1008 Gly Lys Asp Tyr Tyr Val Tyr Thr Cys Arg Lys Asn Tyr Ser Gly Ala aag gac cgc ggc tgc gga aaa gaa atg tct gag aat aaa ttg aac cgg 1056 Lys Asp Arg Gly Cys Gly Lys Glu Met Ser Glu Asn Lys Leu Asn Arg 345 cat gta tgg ggt gaa att ttt aaa ttc atc aca aat cct caa aag tat 1104 45 His Val Trp Gly Glu Ile Phe Lys Phe Ile Thr Asn Pro Gln Lys Tyr gtt tot ttt aaa gag got gaa caa toa aat cac otg tot gat gaa tta 1152 Val Ser Phe Lys Glu Ala Glu Gln Ser Asn His Leu Ser Asp Glu Leu 50 375 gaa ctt att gaa aaa gag ata gag aaa aca aaa aggc cgc aag cgt 1200 Glu Leu Ile Glu Lys Glu Ile Glu Lys Thr Lys Lys Gly Arg Lys Arg 385 55 ctt tta acg cta atc agc cta agc gat gac gat gat tta gac ata gat 1248 Leu Leu Thr Leu Ile Ser Leu Ser Asp Asp Asp Leu Asp Ile Asp 60 gaa atc aaa gca caa att att gaa ctg caa aaa aag caa aat cag ctt 1296 Glu Ile Lys Ala Gln Ile Ile Glu Leu Gln Lys Lys Gln Asn Gln Leu act gaa aag tgt aac aga atc cag tca aaa atg aaa gtc cta gat gat 1344 Thr Glu Lys Cys Asn Arg Ile Gln Ser Lys Met Lys Val Leu Asp Asp 435 440

| | Ser | Tyr | Val | Ser | Lys 245 | Gln | Ala | Gly | Asn | Lys 250 | Ser | Ile | Ile | Lys | Ile 255 | Arg | |
|----|-------------------------|--------------|------------|------------|-----------------|------------|------------|------------|------------|------------------|-------------------|------------|------------|------------|------------------|------------|----|
| 5 | Pro | Glu | Glu | Glu 260 | Gln | Ile | Thr | Val | Thr 265 | Ile | Pro | Ala | Ïle | Val 270 | Pro | Ala | |
| 10 | Glu | Gln | Trp 275 | Asp | Tyr | Ala | Gln | Glu 280 | Leu | Leu | Gly | Gln | Ser 285 | Lys | Arg | Lys | |
| | His | Leu 290 | Ser | Ile | Ser | Pro | His 295 | Asn | Tyr | Leu | Leu | Ser 300 | Gly | Leu | Val | Arg | |
| 15 | Cys 305 | Gly | Lys | Cys | Gly | Asn 310 | Thr | Met | Thr | Gly. | Lys 315 | Lys | Arg | Lys | Ser | His 320 | |
| | Gly | Lys | Asp | Tyr | Tyr 325 | Val | Tyr | Thr | Cys | Arg 330 | Lys | Asn | Tyr | Ser | Gly 335 | Ala | |
| 20 | Lys | Asp | Arg | Gly 340 | Суз | Gly | Lys | Glu | Met 345 | Ser | Glu | Asn | Lys | Leu 350 | Asn | Arg | |
| 25 | His | Val | Trp 355 | Gly | Glu | Ile | Phe | Lys 360 | Phe | Ile | Thr | Asn | Pro 365 | Gln | Lys | Tyr | |
| | Val | Ser 370 | Phe | Lys | Glu | Ala | Glu 375 | Gln | Ser | Asn | His | Leu 380 | Ser | Asp | Glu | Leu | |
| 30 | Glu 385 | Leu | Ile | Glu | Lys | Glu 390 | Ile | Glu | Lys | Thr | Lys 395 | Lys | Gly | Arg | Lys | Arg 400 | |
| | Leu | Ļeu | Thr | Leu | Ile 405 | Ser | Leu | Ser | Asp | Asp 410 | Asp | Asp | Leu | Asp | Ile 415 | Asp | |
| 35 | Glu | Ile | Lys | Ala 420 | Gln | Ile | Ile | Glu | Leu 425 | Gln | Lys | Lys | Gln | Asn 430 | Gln | Leu | |
| 40 | Thr | Glu | Lys 435 | Cys | Asn | Arg | Ile | Gln 440 | Ser | Lys | Met | Lys | Val 445 | Leu | Asp | Asp | |
| | Thr | Ser 450 | Ser | Ser | Glu | Asn | Ala 455 | Leu | Lys | Arg | Ala | Ile 460 | Asp | Tyr | Phe | Gln | |
| 45 | Ser 465 | Ile | Gly | Ala | Asp | Asn 470 | Leu | Thr | Leu | Glu | Asp 475 | Lys | Lys | Thr | Ile | Val 480 | |
| | Asn | Phe | Ile | Val | Lys 485 | Glu | Val | Thr | Ile | Val 490 | Asp | Ser | Asp | Thr | Ile 495 | Tyr | |
| 50 | Ile | Glu | Thr | Tyr 500 | | | | | | | | | | | | | |
| 55 | <210 |)> 58 | | | | | | • | | | | | | | | | |
| | <211 <212 | > 15 > DN | 545 NA | econ | nbina | ıse | | | | | | · | | | | | |
| 60 | <220 <221 |)> .> CE | os | 1542 | | | | | | | | | | • | | | • |
| 65 | <400 atg Met 1 | gaa | aat | tgg Trp | ggt Gly 5 | tac Tyr | gcg Ala | aga Arg | gtg Val | agc Ser 10 | ggt Gly | gag Glu | gaa Glu | cag Gln | caa Gln 15 | aca Thr | 48 |

| 5 . | | | | | | | | | | | | | | | gct Ala | | 96 |
|-----|------------|-------------------|------------|-------------------|------------|------------|-------------------|------------|-------------------|------------|------------|-------------------|------------|-------------------|-------------------|-------------------|-----|
| | | | | | | | | | | | | | | | gtc Val | | 144 |
| 10 | | | | | | | | | | | | | | | aca Thr | | 192 |
| 15 | | | | | | | | | | | | | | | tca Ser | | 240 |
| 20 | | | | | | | | | | | | | | | gga Gly 95 | | 288 |
| 25 | aaa Lys | ctg Leu | ata Ile | gcc Ala 100 | tta Leu | gat Asp | caa Gln | ggc Gly | gtt Val 105 | gac Asp | cca Pro | gac Asp | agc Ser | ctt Leu 110 | ggc Gly | GJÀ aaa | 336 |
| | | | | | | | | | | | | | | | gta Val | | 384 |
| 30 | | | | | | | | | | | | | | | aac Asn | | 432 |
| 35 | | | | | | | | | | | | | | | aaa Lys | | 480 |
| 40 | | | | | | | | | | | | | | | gga Gly 175 | | 528 |
| 45 | | | | | | | | | | | | | | | act Thr | | 576 |
| ,, | | | | | | | | | | | | | | | tca Ser | | 624 |
| 50 | ttt Phe | ggt Gly 210 | ata Ile | gaa Glu | aca Thr | aaa Lys | gtt Val 215 | ctg Leu | aat Asn | tgg Trp | aac Asn | aag Lys 220 | cta Leu | gaa Glu | aaa Lys | tct Ser | 672 |
| 55 | | | | | | | | | | | | | | | aca Thr | cca Pro 240 | 720 |
| 60 | | | | | | | | | | | | | | | | tgg Trp | 768 |
| 65 | | | | | | | | | | | | | | | | GJ À aaa | 816 |
| UJ | | | | | | | | | | | | | | | aag Lys | | 864 |

275 280 285

| | | | 275 | | | | | 280 | | | | | 285 | | | | |
|----|-------------------|----------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------------|-------------------|-------------------|-------|
| 5 | ttt Phe | gac Asp 290 | gag Glu | tgg Trp | aaa Lys | gta Val | aaa Lys 295 | tgg Trp | gga Gly | acc Thr | cac His | gac Asp 300 | gat Asp | gag Glu | gca Ala | atc Ile | 912 |
| 10 | att Ile 305 | acc Thr | tgt Cys | gag Glu | gaa Glu | cat His 310 | gaa Glu | aga Arg | ata Ile | aaa Lys | cag Gln 315 | atg Met | att Ile | cga Arg | gac Asp | aat Asn 320 | 960 |
| | cgc Arg | aat Asn | aat Asn | cga Arg | tgg Trp 325 | gct Ala | gca Ala | aga Arg | gaa Glu | gaa Glu 330 | aac Asn | gaa Glu | gta Val | aac Asn | cca Pro 335 | ttt Phe | 1008 |
| 15 | tct Ser | aat Asn | tta Leu | ctt Leu 340 | aaa Lys | tgt Cys | acc Thr | cat His | tgc Cys 345 | GJ Y Ggc | ggc Gly | tca Ser | atg Met | aca Thr 350 | cgc Arg | cac His | 1056 |
| 20 | gcc Ala | aaa Lys | cgt Arg 355 | gta Val | gat Asp | aag Lys | agt Ser | gga Gly 360 | caa Gln | gct Ala | atc Ile | tat Tyr | tat Tyr 365 | tat Tyr | cag Gln | tgc Cys | 1104 |
| 25 | cga Arg | ttg Leu 370 | tat Tyr | aaa Lys | gct Ala | ggc Gly | aac Asn 375 | tgt Cys | agc Ser | aat Asn | aaa Lys | aat Asn 380 | atg Met | att Ile | tca Ser | tcc Ser | 1152 |
| 30 | aaa Lys 385 | ata Ile | tta Leu | gat Asp | atc Ile | caa Gln 390 | gta Val | atg Met | gat Asp | tta Leu | ttg Leu 395 | gca Ala | caa Gln | gaa Glu | gcc Ala | gaa Glu 400 | 1,200 |
| | cgt Arg | tta Leu | gca Ala | aat Asn | ttg Leu 405 | gtg Val | gaa Glu | aca Thr | gat Asp | gag Glu 410 | ccg Pro | ctt Leu | att Ile | gta Val | gaa Glu 415 | gaa Glu | 1248 |
| 35 | ccc Pro | cca. Pro | gaa Glu | gta Val 420 | aaa Lys | acg Thr | ctg Leu | cgc Arg | gca Ala 425 | tcc Ser | ctg Leu | aat Asn | agt Ser | ctg Leu 430 | gaa Glu | aca Thr | 1296 |
| 40 | ttg Leu | cca Pro | gca Ala 435 | agt Ser | tca Ser | gca Ala | att Ile | gaa Glu 440 | caa Gln | att Ile | aaa Lys | aat Asn | gac Asp 445 | Leu | aaa Lys | gaa Glu | 1344 |
| 45 | cag Gln | att Ile 450 | gcg Ala | atc Ile | gca Ala | cta Leu | gga Gly 455 | gca Ala | acc Thr | aat Asn | aat Asn | gct Ala 460 | tct Ser | aaa Lys | caa Gln | tct Ser | 1392 |
| 50 | ctg Leu 465 | att Ile | gcc Ala | aag Lys | gaa Glu | aga Arg 470 | att Ile | ata Ile | caa Gln | gct Ala | ttt Phe 475 | gct Ala | cat His | aaa Lys | agt Ser | tac Tyr 480 | 1440 |
| | tgg Trp | caa Gln | gga Gly | cta Leu | aac Asn 485 | gct Ala | caa Gln | gat Asp | aaa Lys | cga Arg 490 | gca Ala | atc Ile | ctc Leu | aat [.] Asn | ggt Gly 495 | tgc Cys | 1488 |
| 55 | gta Val | aaa Lys | aaa Lys | atc Ile 500 | tcc Ser | gta Val | gat Asp | ggt Gly | aac Asn 505 | ttt Phe | gtt Val | aca Thr | gct Ala | att Ile 510 | gag Glu | tat Tyr | 1536 |
| 60 | | tac Tyr | tag | | | | | | | | | | | | | | 1545 |
| 65 | <213 <212 | 0> 5: L> 5: 2> PI 3> X: | l 4 RT | recor | mbina | ase | | | | | | • | | | | | |

| | | | | | | | | | | 39 | | | | | | |
|----|------------|-------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | | 0> 5 Glu | | Trp | Gly 5 | Tyr | Ala | Arg | Val | Ser 10 | Gly | Glu | Glu | Gln | Gln 15 | Thr |
| 5 | Asp | Lys | Gly | Ala 20 | Leu | Arg | Lys | Gln | Ile 25 | Glu | Arg | Leu | Arg | Asn 30 | Ala | Gly |
| 10 | Cys | Ser | Lys 35 | Val | Tyr | Trp | Asp | Ile 40 | Gln | Ser | Arg | Thr | Thr 45 | Glu | Val | Arc |
| | | 50 | | | | • | 55 | | | | | 60 | | | Thr | |
| 15 | 65 | | | | | 70 | | | | | 75 | | | | Ser | 80 |
| •• | | | | | 85 | | | | | 90 | | | | | Gly 95 | |
| 20 | | | | 100 | | | | | 105 | | | ÷ | | 110 | Gly | |
| 25 | Glu | Leu | Thr 115 | Ile | Asp | Met | Leu | Leu 120 | Ala | Ala | Ala | Lys | Phe 125 | Glu | Val | Arg |
| | Met | Val 130 | Thr | Glu | Arg | Leu | Lys 135 | Ser | Glu | Arg | Arg | His 140 | Arg | Val | Asn | Gln |
| 30 | Gly 145 | Lys | Ser | His | Arg | Val 150 | Ala | Pro | Leu | Gly | Tyr 155 | Arg | Lys | Asp | Lys | Asp 160 |
| • | Lys | Tyr | Ile | Arg | Asp 165 | Arg | Ser | Pro | Суз | Val 170 | Cys | Leu | Leu | Glu | Gly 175 | Arg |
| 35 | Arg | Glu | Leu | Thr 180 | Val | Ser | Asp | Leu | Ala 185 | Gln | Tyr | Ile | Phe | His 190 | Thr | Phe |
| 40 | Phe | Glu | Cys 195 | Gly | Ser | Val | Ala | Ala 200 | Thr | Val | Arg | Lys | Leu 205 | His | Ser | Asp |
| | Phe | Gly 210 | Ile | Glu | Thr | Lys | Val 215 | Leu | Asn | Trp | Asn | Lys 220 | Leu | Glu | Lys | Ser |
| 45 | Ser 225 | Arg | Ile | Val | Gly | Asp 230 | Asp | Asp | Leu | Asp | Lys 235 | Ile | Ala | Phe | Thr | Pro 240 |
| | Asn | Lys | Thr | Asn | His 245 | Pro | Leu | Arg | Tyr | Pro 250 | Trp | Ser | Gly | Leu | Arg 255 | Trp |
| 50 | Ser | Ile | Pro | Gly 260 | Leu | Lys | Ala | Leu | Leu 265 | Val | Asn | Pro | Val | Tyr 270 | Ala | Gly |
| 55 | Gly | Leu | Pro 275 | Phė | Asp | Thr | Tyr | Val 280 | Lys | Ser | Lys | Gly | Lys 285 | Arg | Lys | His |
| | Phe | Asp 290 | Glu | Trp | Lys | Val | Lys 295 | Trp | Gly | Thr | His | Asp 300 | Asp | Glu | Ala | Ile |
| 60 | Ile 305 | Thr | Cys | Glu | Glu | His 310 | Glu | Arg | Ile | Lys | Gln 315 | Met | Ile | Arg | Asp | Asn 320 |
| | Arg | Asn | Asn | Arg | Trp 325 | Ala | Ala | Arg | Glu | Glu 330 | Asn | Glu | Val | Asn | Pro 335 | Phe |
| 55 | Ser | Asn | Leu | Leu 340 | Lys | Cys | Thr | His | Cys 345 | Gly | Gly | Ser | Met | Thr 350 | Arg | His |

| | Ala | Lys | Arg 355 | Val | Asp | Lys | Ser | Gly 360 | Gln | Ala | Ile | Tyr | Tyr 365 | Tyr | Gln | Cys | |
|----|------------------|-------------------------|------------------|------------------|-----------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|-----|
| 5 | Arg | Leu 370 | Tyr | ГЛS | Ala | Gly | Asn 375 | Суѕ | Ser | Asn | ГÀЗ | Asn 380 | Met | Ile | Ser | Ser | |
| | Lys 385 | Ile | Leu | Asp | Ile | Gln 390 | Val | Met | Asp | Leu | Leu 395 | Ala | Gln | Glu | Ala | Glu 400 | |
| 10 | Arg | Leu | Ala | Asn | Leu 405 | Val | Glu | Thr | Asp | Glu 410 | Pro | Leu | Ile | Val | Glu 415 | Glu | |
| 15 | Pro | Pro | Glu | Val 420 | Lys | Thr | Leu | Arg | Ala 425 | Ser | Leu | Asn | Ser | Leu 430 | Glu | Thr | |
| 10 | Leu | Pro | Ala 435 | Ser | Ser | Ala | Ile | Glu 440 | Gln | Ile | Lys | Asn | Asp 445 | Leu | Lys | Glu | |
| 20 | Gln | Ile 450 | Ala | Ile | Ala | Leu | Gly 455 | Ala | Thr | Asn | Asn | Ala 460 | Ser | Lys | Gln | Ser | |
| | Leu 465 | Ile | Ala | Lys | Glu | Arg 470 | Ile | Ile | Gln | Ala | Phe 475 | Ala | His | Буs | Ser | Tyr 480 | |
| 25 | Trp | Gln | Gly | Leu | Asn 485 | Ala | Gln | Asp | Lys | Arg 490 | Ala | Ile | Leu | Asn | Gly 495 | Cys | |
| 30 | Val | Lys | Lys | Ile 500 | Ser | Val | Asp | Gly | Asn 505 | Phe | Val | Thr | Ala | Ile 510 | Glu | Tyr | |
| | Arg | Tyr | | | | | | | | | | | | | | | |
| 35 | <211 <212 |)> 6(L> 2: 2> Di | L24 NA | | - m | 4451 | | | | | | | | | | | |
| 40 | <220 <221 |)> L> CI | | | n Tn | 4431 | | | | | | | | | | | |
| 45 | atg |)> 60 tca Ser | agg | act Thr | tca Ser 5 | aga Arg | att Ile | aca Thr | gca Ala | ctt Leu 10 | tac Tyr | gag Glu | cgt Arg | ttg Leu | tca Ser 15 | aga Arg | 48 |
| 50 | gat Asp | gat Asp | gac Asp | ctt Leu 20 | act Thr | ggc Gly | gag Glu | agt Ser | aat Asn 25 | tct Ser | att Ile | acc Thr | aat Asn | caa Gln 30 | aag Lys | aaa Lys | 96 |
| 55 | tac Tyr | ctc Leu | gaa Glu 35 | gat Asp | tat Tyr | gcc Ala | cgt Arg | agg Arg 40 | aat Asn | ggt Gly | ttt Phe | gag Glu | aac Asn 45 | att Ile | cgc Arg | cat His | 144 |
| 60 | ttt Phe | acc Thr 50 | gat Asp | gac Asp | gga Gly | ttt Phe | tcg Ser 55 | ggt Gly | gta Val | aat Asn | ttc Phe | aat Asn 60 | cgc Arg | cct Pro | ggc Gly | ttt Phe | 192 |
| 65 | caa Gln 65 | tct Ser | ctg Leu | ata Ile | aaa Lys | gaa Glu 70 | gtt Val | gaa Glu | gca Ala | gga Gly | aat Asn 75 | gta Val | gaa Glu | acc Thr | ttg Leu | att Ile 80 | 240 |
| | gtt Val | aag Lys | gat Asp | atg Met | agc Ser | cga Ara | ttg Leu | ggg Glv | cga Arg | aat | tat | ctg | caa | gta Val | ggt Glv | ttt Phe | 288 |

41 85 90 tat acg gaa gtt ctg ttt cca cag aaa aat gtc cgt ttc ctt gca att Tyr Thr Glu Val Leu Phe Pro Gln Lys Asn Val Arg Phe Leu Ala Ile 105 aac aac agt att gac agt aac aac gct tcg gat aat gac ttt gct ccg 384 Asn Asn Ser Ile Asp Ser Asn Asn Ala Ser Asp Asn Asp Phe Ala Pro 10 ttt ttg aat att atg aac gaa tgg tat gcc aaa gac aca agc aac aaa 432 Phe Leu Asn Ile Met Asn Glu Trp Tyr Ala Lys Asp Thr Ser Asn Lys 135 atc aag gct ata ttc gat gcc cgt atg aaa gac gga aag cgt tgt agc Ile Lys Ala Ile Phe Asp Ala Arg Met Lys Asp Gly Lys Arg Cys Ser 145 ggt tot atc cot tat ggg tat aac cga ctg ccg agc gac aaa caa acg 528 Gly Ser Ile Pro Tyr Gly Tyr Asn Arg Leu Pro Ser Asp Lys Gln Thr 165 ctt gtg gtt gac cet gtg get teg gaa gtg gta aag egt ate ttt act Leu Val Val Asp Pro Val Ala Ser Glu Val Val Lys Arg Ile Phe Thr 25 185 ctt gcc aat gat ggc aaa agt aca agg gca atc gca gaa ata ctg acc 624 Leu Ala Asn Asp Gly Lys Ser Thr Arg Ala Ile Ala Glu Ile Leu Thr 30 gaa gaa aaa gtt tta acc cct gcg gca tac gca aag gaa tac cac ccc 672 Glu Glu Lys Val Leu Thr Pro Ala Ala Tyr Ala Lys Glu Tyr His Pro 215 gaa cag tac aac ggc aac aag ttc aca aac cct tat ctt tgg gca atg 720 Glu Gln Tyr Asn Gly Asn Lys Phe Thr Asn Pro Tyr Leu Trp Ala Met 230 235 tca acg ata aga aat att tta ggc agg cag gaa tat ctc ggt cac acc 768 Ser Thr Ile Arg Asn Ile Leu Gly Arg Gln Glu Tyr Leu Gly His Thr 245 gtt ttg cga aag tcg gta agc aca aat ttc aaa ctt cac aag aga aaa 816 Val Leu Arg Lys Ser Val Ser Thr Asn Phe Lys Leu His Lys Arg Lys 45 260 agc aca gac gaa gaa cag tat gta ttt ccg aat aca cac gag cct Ser Thr Asp Glu Glu Glu Gln Tyr Val Phe Pro Asn Thr His Glu Pro 285 50 atc ata tcg cag gaa ctt tgg gac agc gtt caa aaa cgc aga agc aga 912 Ile Ile Ser Gln Glu Leu Trp Asp Ser Val Gln Lys Arg Arg Ser Arg 300 55 gta aat cgt gcc tcg gct tgg gga acg cac agc aac cgt tta agc gga 960 Val Asn Arg Ala Ser Ala Trp Gly Thr His Ser Asn Arg Leu Ser Gly 310 tat ttg tac tgt gcc gat tgc gga aga aga atg act ttg cag aca cat 1008 60 Tyr Leu Tyr Cys Ala Asp Cys Gly Arg Arg Met Thr Leu Gln Thr His 330 tac agc aaa aaa gac ggt tot gtg cag tat tot tac cgt tgc ggt ggg 1056 Tyr Ser Lys Lys Asp Gly Ser Val Gln Tyr Ser Tyr Arg Cys Gly Gly 65 340 tat gca agc aga gtg aac agt tgt acc agt cat tcg att agt acc gat 1104

WO 02/38613

Tyr Ala Ser Arg Val Asn Ser Cys Thr Ser His Ser Ile Ser Thr Asp

aat gtt gaa gcc ttg ata tta tca tct gtc aaa cgc ttt tca agg ttt 1152
Asn Val Glu Ala Leu Ile Leu Ser Ser Val Lys Arg 380

PCT/EP01/12975

gtt ctg aat gat gaa caa gca ttt gct ttg gaa ctg caa tct ctt tgg 1200
Val Leu Asn Asp Glu Gln Ala Phe Ala Leu Glu Leu Gln Ser Leu Trp
385 390 395 400

15

35

55

Asn Glu Lys Gln Glu Glu Lys Pro Lys His Asn Gln Ser Glu Leu Gln 405 410 415

cgc tgt cag aaa cgc tat gac gaa ctc tct acc ctt gtt cgt ggc ttg
Arg Cys Gln Lys Arg Tyr Asp Glu Leu Ser Thr Leu Val Arg Gly Leu
420 425 430

20 tat gaa aat ctt atg tcg gga tta ctg ccc gaa aga cag tat aag caa 1344 Tyr Glu Asn Leu Met Ser Gly Leu Leu Pro Glu Arg Gln Tyr Lys Gln . 435 440 445

ctg atg aaa cag tat gat gac gag cag gca gag ttg gaa acg aaa atg 1392

Leu Met Lys Gln Tyr Asp Asp Glu Gln Ala Glu Leu Glu Thr Lys Met
450

460

gaa acg atg aaa aca gaa ctt gcc gaa gaa aaa gta agt tcc gtt gat
Glu Thr Met Lys Thr Glu Leu Ala Glu Glu Lys Val Ser Ser Val Asp
465
470
480

att aag cat ttc att tcg ctg ata cgc aag tgt aaa aat cct acg gaa I488 Ile Lys His Phe Ile Ser Leu Ile Arg Lys Cys Lys Asn Pro Thr Glu 485 490 495

atc tcc gat aca atg ttt aat gaa ctt gtt gat aag ata gtg gtt tat 1536

Ile Ser Asp Thr Met Phe Asn Glu Leu Val Asp Lys Ile Val Val Tyr

500 505 510

gaa gca gag ggt gtg gga aaa gca cga aca caa aag gtc gat att tat 1584 Glu Ala Glu Gly Val Gly Lys Ala Arg Thr Gln Lys Val Asp Ile Tyr 515

ttt aac tat gtc ggt caa gtg gat att gcc tat acc gaa gaa ctt 1632
45 Phe Asn Tyr Val Gly Gln Val Asp Ile Ala Tyr Thr Glu Glu Glu Leu
530 540

gcc gag ata gaa aca cag aaa gag cag gag gaa cag caa cgc ttg gca 1680
Ala Glu Ile Glu Thr Gln Lys Glu Glu Glu Glu Gln Arg Leu Ala
50 545 550 555 560

aga cag cgc aag cgt gaa aaa gcc tac cga gaa aag cga aag gca cag
Arg Gln Arg Lys Arg Glu Lys Ala Tyr Arg Glu Lys Arg Lys Ala Gln
565
570
575

aaa atc gct gaa aac ggt ggc gaa atc gtt aag aca aag gtt tgc cct 1776 Lys Ile Ala Glu Asn Gly Gly Glu Ile Val Lys Thr Lys Val Cys Pro 580 585

60 cat tgc aac aaa gag ttt atc ccg aca agc aac cga cag gtg ttc tgt
His Cys Asn Lys Glu Phe Ile Pro Thr Ser Asn Arg Gln Val Phe Cys
595 600 605

tcc aaa gag tgc tgc tat caa gca agg caa gac aaa aag aaa aca gac 1872 Ser Lys Glu Cys Cys Tyr Gln Ala Arg Gln Asp Lys Lys Thr Asp 610 615 620

Leu Val Val Asp Pro Val Ala Ser Glu Val Val Lys Arg Ile Phe Thr

180

| | Leu | ı Ala | Asn 195 | Asp | Gly | Lys | Ser | Thr 200 | Arg | Ala | lle | Ala | Glu 205 | Ile | Leu | Thr |
|----|------------|------------|------------|------------|------------|------------|------------|-------------|------------|------------|------------|-------------------------|------------|------------|------------|------------|
| 5 | Glu | Glu 210 | Lys | Val | Leu | Thr | Pro 215 | Ala | Ala | Tyr | Ala | Lys 220 | | Tyr | His | Pro |
| | Glu 225 | Gln | Tyr | Asn | Gly | Asn 230 | Lys | Phe | Thr | Asn | Pro 235 | | Leu | Trp | Ala | Met 240 |
| 10 | Ser | Thr | Ile | Arg | Asn 245 | Ile | Leu | Gly | Arg | Gln 250 | Glu | Tyr | Leu | Gly | His 255 | |
| 15 | Val | Leu | Arg | Lys 260 | Ser | Val | Ser | Thr | Asn 265 | Phe | Lys | Leu | His | Lys 270 | | Lys |
| | Ser | Thr | Asp 275 | Glu | Glu | Glu | Gln | Tyr 280 | Val | Phe | Pro | Asn | Thr 285 | His | Glu | Pro |
| 20 | Ile | 11e 290 | Ser | Gln | Glu | Leu | Trp 295 | Asp | Ser | Val | Gln | Lys 300 | | Arg | Ser | Arg |
| | Val 305 | Asn | Arg | Ala | Ser | Ala 310 | Trp | Gly | Thr | His | Ser 315 | Asn | Arg | Leu | Ser | Gly 320 |
| 25 | Tyr | Leu | Tyr | Cys | Ala 325 | Asp | Суз | Gly | Arg | Arg 330 | Met | Thr | Leu | Gln | Thr 335 | His |
| 30 | Tyr | Ser | Lys | Lys 340 | Asp | Gly | Ser | Val | Gln 345 | Tyr | Ser | Tyr | Arg | Cys 350 | Gly | Gly |
| | Tyr | Ala | Ser 355 | Arg | Val | Asn | Ser | Cys 360 | Thr | Ser | His | Ser | Ile 365 | Ser | Thr | Asp |
| 35 | Asn | Val 370 | Glu | Ala | Leu | Ile | Leu 375 | Ser | Ser | Val | Lys | Arg 380 | Phe | Ser | Arg | Phe |
| | Val 385 | Leu | Asn | Asp | Glu | Gln 390 | Ala | Phe | Ala | Leu | Glu 395 | Leu | Gln | Ser | Leu | Trp 400 |
| 40 | Asn | Glu | Lys | Gln | Glu 405 | Glu | Lys | Pro | Lys | His 410 | Asn | Gln | Ser | Glu | Leu 415 | Gln |
| 45 | Arg | Суѕ | Gln | Lys 420 | Arg | Tyr | Asp | Gl u | Leu 425 | Ser | Thr | Leu | Val | Arg 430 | Gly | Leu |
| | Tyr | Glu | Asn 435 | Leu | Met | Ser | Gly | Leu 440 | Leu | Pro | Glu | Arg | Gln 445 | Tyr | Lys | Gln |
| 50 | Leu | Met 450 | Lys | Gln | Tyr | Asp | Asp 455 | Glu | Gln | Ala | Glu | Leu [.] 460 | Glu | Thr | Lys | Met |
| | Glu 465 | Thr | Met | Lys | Thr | Glu 470 | Leu | Ala | Glu | Glu | Lys 475 | Val | Ser | Ser | Val | Asp 480 |
| 55 | Ile | Lys | His | Phe | Ile 485 | Ser | Leu | Ile | Arg | Lys 490 | Cys | Lys | Asn | Pro | Thr 495 | Glu |
| 60 | Ile | Ser | Asp | Thr 500 | Met | Phe | Asn | Glu | Leu 505 | Val | Asp | Lys | Ile | Val 510 | Val | Tyr |
| | Glu | Ala | Glu 515 | Gly | Val | Gly | Lys | Ala 520 | Arg | Thr | Gln | Lys | Val 525 | Asp | Ile | Tyr |
| 65 | Phe | Asn 530 | Tyr | Val | Gly | Gln | Val 535 | Asp | Ile | Ala | Tyr | Thr 540 | Glu | Glu | Glu | Leu |
| | Ala | Glu | Ile | Glu | Thr | Gln | Lys | Glu | Gln | Glu | Glu | Gln | Gln | Dra | T.em | Δla |

WO 02/38613

| 5 | agc Ser | ttg Leu | aat Asn | atc Ile 100 | cct Pro | gca Ala | aac Asn | ttg Leu | gat Asp 105 | gga Gly | ctg Leu | aag Lys | acg Thr | gct Ala 110 | gag Glu | gaa Glu | 336 |
|----|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|------|
| | gaa Glu | gct Ala | tat Tyr 115 | gaa Glu | tta Leu | ggt Gly | aaa Lys | tta Leu 120 | atc Ile | gct Ala | cgg Arg | aaa Lys | acc Thr 125 | ttt Phe | gaa Glu | tgg Trp | 384 |
| 10 | aat Asn | gat Asp 130 | ьys | tat Tyr | tta Leu | ggc Gly | aaa Lys 135 | gaa Glu | gcc Ala | act Thr | aaa Lys | aaa Lys 140 | gat Asp | tca Ser | caa Gln | aca Thr | 432 |
| 15 | ata Ile 145 | ggt Gly | gat Asp | tta Leu | cta Leu | gaa Glu 150 | aaa Lys | ttt Phe | gca Ala | gaa Glu | gag Glu 155 | tat Tyr | ttt Phe | aaa Lys | acc Thr | cat His 160 | 480 |
| 20 | aaa Lys | cgc Arg | acc Thr | act Thr | aaa Lys 165 | agc Ser | gaa Glu | cat His | acc Thr | ttt Phe 170 | ttt Phe | tac Tyr | tat Tyr | ttt Phe | tcc Ser 175 | cgc Arg | 528 |
| 25 | acc Thr | caa Gln | cga Arg | tat Tyr 180 | acc Thr | aat Asn | tcc Ser | aaa Lys | gat Asp 185 | tta Leu | gca Ala | acg Thr | gcg Ala | gaa Glu 190 | aat Asn | ctc Leu | 576 |
| | atc Ile | aat Asn | tca Ser 195 | att Ile | gag Glu | caa Gln | atc Ile | gat Asp 200 | aaa Lys | gaa Glu | tgg Trp | gcg Ala | aga Arg 205 | tat Tyr | aat Asn | gcc Ala | 624 |
| 30 | gcc Ala | aga Arg 210 | gcc Ala | ata Ile | tca Ser | gct Ala | ttt Phe 215 | tgc Cys | ata Ile | aca Thr | ttc Phe | aat Asn 220 | ata Ile | gaa Glu | att Ile | gat Asp | 672 |
| 35 | ttg Leu 225 | tcc Ser | cag Gln | tat Tyr | tcc Ser | aaa Lys 230 | atg Met | cct Pro | gat Asp | cgc Arg | aat Asn 235 | tcg Ser | cgc Arg | aac Asn | atc Ile | ccc Pro 240 | 720 |
| 40 | aca Thr | gat Asp | gca Ala | gaa Glu | ata Ile 245 | cta Leu | tca Ser | gga Gly | att Ile | acc Thr 250 | aaa Lys | ttt Phe | gaa Glu | gac Asp | tat Tyr 255 | cta Leu | 768 |
| 45 | gtt Val | acc Thr | aga Arg | gga Gly 260 | aat Asn | caa Gln | gtt Val | aat Asn | gaa Glu 265 | gat Asp | gta Val | aaa Lys | gat Asp | agc Ser 270 | tgg Trp | caa Gln | 816 |
| | ctt Leu | tgg Trp | cgc Arg 275 | tgg Trp | aca Thr | tat Tyr | gga Gly | atg Met 280 | tta Leu | gca Ala | gtt Val | ttt Phe | ggt Gly 285 | tta Leu | cgc Arg | ccc Pro | 864 |
| 50 | agg Arg | gaa Glu 290 | att Ile | ttt Phe | att Ile | aac Asn | cct Pro 295 | aat Asn | att Ile | gat Asp | tgg Trp | tgg Trp 300 | tta Leu | agc Ser | aaa Lys | gag Glu | 912 |
| 55 | aat Asn 305 | ata Ile | gac Asp | ctc Leu | aca Thr | tgg Trp 310 | aaa Lys | gta Val | gac Asp | aaa Lys | gaa Glu 315 | tgt Cys | aaa Lys | act Thr | ggt Gly | gaa Glu 320 | 960 |
| 60 | aga Arg | caa Gln | gca Ala | tta Leu | ccc Pro 325 | tta Leu | cat His | aaa Lys | gaa Glu | tgg Trp 330 | att Ile | gat Asp | gag Glu | ttt Phe | gat Asp 335 | tta Leu | 1008 |
| 65 | aga Arg | aat Asn | ccg Pro | aaa Lys 340 | tat Tyr | tta Leu | gaa Glu | atg Met | ctg Leu 345 | gca Ala | aca Thr | gca Ala | att Ile | agt Ser 350 | aaa Lys | aaa Lys | 1056 |
| | gat Asp | aaa Lys | aca Thr | aat Asn | cat His | gct Ala | gaa Glu | ata Ile | aca Thr | gcc Ala | tta Leu | act Thr | cag Gln | cgt Arg | att Ile | agt Ser | 1104 |

| | | | 355 | | | | | 360 | | 4/ | | | 365 | | | | |
|------|--------------------------------|---------------------------------------|--|--|--------------------------------|-------------------------|--------------------------------|--|--------------------------------|--------------------------------|--------------------------------|----------------------------|---------------------------------------|---|---------------------------------------|---------------------------------------|------|
| 5 | tgg Trp | tgg Trp 370 | ttt Phe | cgg Arg | aaa Lys | gtc Val | gaa Glu 375 | tta Leu | gat Asp | ttt Phe | aaa Lys | ccc Pro 380 | tat Tyr | gat Asp | tta Leu | cgt Arg | 1152 |
| 10 | cac His 385 | gcc Ala | tgg Trp | gca Ala | atc Ile | aga Arg 390 | gcg Ala | cat His | att Ile | tta Leu | ggc Gly 395 | ata Ile | cca Pro | atc Ile | aaa Lys | gcg Ala 400 | 1200 |
| 10 | gcg Ala | gct Ala | gat Asp | aat Asn | ttg Leu 405 | gly ggg | cat His | agt Ser | atg Met | cag Gln 410 | gtt Val | cat His | aca Thr | caa Gln | acc Thr 415 | tat Tyr | 1248 |
| 15 | cag Gln | cgc Arg | tgg Trp | ttc Phe 420 | tcg Ser | cta Leu | gat Asp | atg Met | cgg Arg 425 | aag Lys | tta Leu | gcg Ala | att Ile | aat Asn 430 | cag Gln | gct Ala | 1296 |
| . 20 | ttg Leu | act Thr | aag Lys 435 | agg Arg | aat Asn | gaa Glu | ttt Phe | gag Glu 440 | gtg Val | att Ile | agg Arg | gag Glu | gag Glu 445 | aat Asn | gct Ala | aaa Lys | 1344 |
| 25 | ttg Leu | cag Gln 450 | ata Ile | gaa Glu | aat Asn | gaa Glu | agg Arg 455 | ttg Leu | agg Arg | atg Met | gaa Glu | att Ile 460 | gag Glu | aag Lys | tta Leu | aag Lys | 1392 |
| 30 | atg Met 465 | gaa Glu | ata Ile | gct Ala | tat Tyr | aag Lys 470 | aat Asn | agt Ser | tgaç | 3 | | | | | | | 1420 |
| 35 | <211 <212 | 0> 63 l> 47 2> PF 3> Xi | 72 RT | recon | nbina | ıse | | | | | | | | | | | |
| 40 | |)> 63 Gln | | Gln | Gly 5 | Gln | Asp | Lys | Tyr | Gln 10 | Gln | Ala | Phe | Ala | Asp 15 | Leu | |
| | Glu | Pro | Leu | Ser 20 | Ser | Thr | Asp | Gly | Ser | - | . | Glu | Ser | Ser | Leu | Gln | |
| 45 | Ala | Gln | | | | | | | 25 | Phe | теп | GLY | DCI | 30 | | 02 | |
| | | | 35 | Gln | Arg | Glu | His | Met 40 | 25 | | | _ | | | | | |
| | Asp | | 35 | | Arg Leu | | | 40 | 25 Arg | Thr | Lys | Val | Leu 45 | Gln | Asp | Leu | |
| 50 | | Lys 50 | 35 Val | Asn | | Arg | Leu 55 | 40 Lys | 25 Arg Ser | Thr Ala | Lys Lys | Val Thr 60 | Leu 45 Lys | Gln Val | Asp Ser | Leu Val | |
| | Arg | Lys 50 Glu | 35 Val Ser | Asn Asn | Leu Gly | Arg Ser 70 | Leu 55 Leu | 40 Lys Gln | 25 Arg Ser Leu | Thr Ala Arg | Lys Lys Ala 75 | Val Thr 60 Thr | Leu 45 Lys Leu | Gln Val Pro | Asp Ser Ile | Leu Val Lys 80 | · |
| 50 | Arg 65 .Pro | Lys 50 Glu Gly | 35 Val Ser Asp | Asn Asn Lys | Leu Gly Asp | Arg Ser 70 Thr | Leu 55 Leu Asn | 40 Lys Gln Gly | 25 Arg Ser Leu Thr | Thr Ala Arg Gly 90 | Lys Lys Ala 75 Arg | Val Thr 60 Thr | Leu 45 Lys Leu Gln | Gln Val Pro Tyr | Asp Ser Ile Asn 95 | Leu Val Lys 80 Leu | |
| | Arg 65 Pro | Lys 50 Glu Gly Leu | 35 Val Ser Asp | Asn Asn Lys | Leu Gly Asp 85 | Arg Ser 70 Thr | Leu 55 Leu Asn Asn | 40 Lys Gln Gly Leu | 25 Arg Ser Leu Thr Asp 105 | Thr Ala Arg Gly 90 Gly | Lys Lys Ala 75 Arg | Val Thr 60 Thr Lys | Leu 45 Lys Leu Gln | Gln Val Pro Tyr Ala | Asp Ser Ile Asn 95 Glu | Leu Val Lys 80 Leu Glu | |
| 55 | Arg 65 Pro Ser Glu | Lys 50 Glu Gly Leu Ala | 35 Val Ser Asp Asn Tyr 115 | Asn Asn Lys Ile 100 Glu | Leu Gly Asp 85 Pro | Arg Ser 70 Thr Ala Gly | Leu 55 Leu Asn Asn | 40 Lys Gln Gly Leu Leu 120 | 25 Arg Ser Leu Thr Asp 105 Ile | Thr Ala Arg Gly 90 Gly Ala | Lys Lys Ala 75 Arg Leu Arg | Val Thr 60 Thr Lys Lys Lys | Leu 45 Lys Leu Gln Thr | Gln Val Pro Tyr Ala 110 Phe | Asp Ser Ile Asn 95 Glu | Leu Val Lys 80 Leu Glu Trp | |

T/EP01/12975

| | v | VO 02 | /3861 | 3 | | | | | | | | | | | | PC |
|-----|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | Lys | Arg | Thr | Thr | Lys | Ser | Glu | His | Thr | 48 Phe | Phe | Tyr | Tyr | Phe | Ser | Arg |
| - 5 | | | | Tyr | 165 | | | | Asp | 170 | | | | Glu | 175 | |
| . , | Ile | Asn | Ser | 180 Ile | Glu | Gln | Ile | Asp | 185 Lys | Glu | Trp | Ala | Ara | 190 Tyr | Asn | Ala |
| 10 | | | 195 | | | | | 200 | | | | | 205 | Glu | | |
| | | 210 | | | | | 215 | | | | | 220 | | | | |
| 15 | ьец 225 | Ser | Gln | Tyr | Ser | Lys 230 | Met | Pro | Asp | Arg | Asn 235 | Ser | Arg | Asn | Ile | Pro 240 |
| • | Thr | Asp | Ala | Glu | Ile 245 | Leu | Ser | Gly | Ile | Thr 250 | Lys | Phe | Glu | Asp | Tyr 255 | Leu |
| 20 | Val | Thr | Arg | Gly 260 | Asn | Gln | Val | Asn | Glu 265 | Asp | Val | Lys | Asp | Ser 270 | Trp | Gln |
| | Leu | Trp | Arg 275 | Trp | Thr | Tyr | Gly | Met 280 | Leu | Ala | Val | Phe | Gly 285 | Leu | Arg | Pro |
| 25 | Arg | Glu 290 | Ile | Phe | Ile | Asn | Pro 295 | Asn | Ile | Asp | Trp | Trp 300 | Leu | Ser | Lys | Glu |
| 30 | Asn 305 | Ile | Asp | Leu | Thr | Trp 310 | Lys | Val | Asp | Lys | Glu 315 | Суз | Lys | Thr | Gly | Glu 320 |
| | Arg | Gln | Ala | Leu | Pro 325 | Leu | His | Lys | Glu | Trp 330 | Ile | Asp | Glu | Phe | Asp 335 | Leu |
| 35 | Arg | Asn | Pro | Lys 340 | Tyr | Leu | Glu | Met | Leu 345 | Ala | Thr | Ala | Ile | Ser 350 | Lys | Lys |
| | Asp | Lys | Thr 355 | Asn | His | Ala | Glu | Ile 360 | Thr | Ala | Leu | Thr | Gln 365 | Arg | Ile | Ser |
| 40 | Trp | Trp 370 | Phe | Arg | Lys | Val | Glu 375 | Leu | Asp | Phe | Lys | Pro 380 | Tyr | Asp | Leu | Arg |
| 45 | His 385 | Ala | Trp | Ala | Ile | Arg 390 | Ala | His | Ile | Leu | Gly 395 | Ile | Pro | Ile | Lys | Ala 400 |
| | Ala | Ala | Asp | Asn | Leu 405 | Gly | His | Ser | Met | Gln 410 | Val | His | Thr | Gln | Thr 415 | Tyr |
| 50 | Gln | Arg | Trp | Phe 420 | Ser | Leu | Asp | Met | Arg 425 | Lys | Leu | Ala | Ile | Asn 430 | Gln | Ala |
| | Leu | Thr | Lys 435 | Arg | Asn | Glu | Phe | Glu 440 | Val | Ile | Arg | Glu | Glu 445 | Asn | Ala | Lys |

Leu Gln Ile Glu Asn Glu Arg Leu Arg Met Glu Ile Glu Lys Leu Lys 450 455 460 Met Glu Ile Ala Tyr Lys Asn Ser 465 470 60

<210> 64

PCT/EP01/12975

49 <220> <221> CDS <222> (1)..(1005) <220> <223> Description of Artificial Sequence: vector pBS-SSV3 <400> 64

10 atg acg aaa gat aag acg cgt tat aaa tac ggg gat tat att tta cgc Met Thr Lys Asp Lys Thr Arg Tyr Lys Tyr Gly Asp Tyr Ile Leu Arg

gag agg aaa ggg cgg tat tat gtt tac aag cta gag tat gaa aac ggt 15 Glu Arg Lys Gly Arg Tyr Tyr Val Tyr Lys Leu Glu Tyr Glu Asn Gly

gag gta aaa gag cgt tac gtg ggt cct tta gct gac gtc gtt gaa tca 144 Glu Val Lys Glu Arg Tyr Val Gly Pro Leu Ala Asp Val Val Glu Ser 20

tat cta aaa atg aaa tta ggg gtc gta ggg gat act ccc cta caa gcg 192 Tyr Leu Lys Met Lys Leu Gly Val Val Gly Asp Thr Pro Leu Gln Ala

gat eee eee ggt tte gag eee ggg aca age gga age ggt ggt gga aaa 240 Asp Pro Pro Gly Phe Glu Pro Gly Thr Ser Gly Ser Gly Gly Lys

gag gga act gaa cga cgt aaa ata gcg ttg gtt gcc aat ttg cgc caa 288 Glu Gly Thr Glu Arg Arg Lys Ile Ala Leu Val Ala Asn Leu Arg Gln

tac gcg acg gac ggc aac ata aag gcg ttc tac aac tat ctc atg aac 336 Tyr Ala Thr Asp Gly Asn Ile Lys Ala Phe Tyr Asn Tyr Leu Met Asn

gaa agg ggg ata agc gaa aaa act gca aag gac tac atc aat gct ata 384 Glu Arg Gly Ile Ser Glu Lys Thr Ala Lys Asp Tyr Ile Asn Ala Ile 40 120

45

65

tca aag ccg tat aaa gag acg aga gac gca cag aag gct tac cga ctc 432 Ser Lys Pro Tyr Lys Glu Thr Arg Asp Ala Gln Lys Ala Tyr Arg Leu

ttt gca cgt ttc tta gcg tca cgc aat atc ata cat gat gaa ttt gcg Phe Ala Arg Phe Leu Ala Ser Arg Asn Ile Ile His Asp Glu Phe Ala

50 gat aaa ata ttg aaa gcg gta aag gtg aag gcg aac gct gat atc 528 Asp Lys Ile Leu Lys Ala Val Lys Val Lys Lys Ala Asn Ala Asp Ile

tac att cca acg ttg gaa gag ata aaa agg acg tta caa tta gca aaa 576 Tyr Ile Pro Thr Leu Glu Glu Ile Lys Arg Thr Leu Gln Leu Ala Lys 190

gac tat agc gaa aac gtc tac ttc atc tac cgt atc gct ctc gag tcg 624 Asp Tyr Ser Glu Asn Val Tyr Phe Ile Tyr Arg Ile Ala Leu Glu Ser 60 200

gge gtt agg etg age gaa ata etg aaa gtg etg aag gaa eee gaa agg 672 Gly Val Arg Leu Ser Glu Ile Leu Lys Val Leu Lys Glu Pro Glu Arg 215

gac att tgc ggt aac gac gtc tgt tat tat ccg ctt agt tgg act agg 720 Asp Ile Cys Gly Asn Asp Val Cys Tyr Tyr Pro Leu Ser Trp Thr Arg

PCT/EP01/12975

WO 02/38613

| | Tyr | Ile | Pro | Thr 180 | Leu | Glu | Glu | Ile | Lys 185 | Arg | Thr | Leu | Gln | Leu 190 | Ala | Lys | |
|----|------------------|----------------------------------|------------------|------------------|-----------------|------------------|------------------|------------------|------------------|------------------------|------------------|------------------|------------------|------------------|------------------|------------------|-----|
| 5 | Asp | Tyr | Ser 195 | Glu | Asn | Val | Tyr | Phe 200 | Ile | Tyr | Arg | Ile | Ala 205 | Leu | Glu | Ser | |
| 10 | Gly | Val 210 | Arg | Leu | Ser | Glu | Ile 215 | Leu | Lys | Val | Leu | Lys 220 | Glu | Pro | Glu | Arg | |
| | Asp 225 | Ile | Cys | Gly | Asn | Asp 230 | Val | Суѕ | Tyr | Tyr | Pro 235 | Leu | Ser | Trp | Thr | Arg 240 | |
| 15 | Gly | Tyr | Lys | Gly | Val 245 | Phe | Tyr | Val | Phe | His 250 | Ile | Thr | Pro | Leu | Lys 255 | Arg | |
| | Val | Glu | Val | Thr 260 | Lys | Trp | Ala | Ile | Ala 265 | Asp | Phe | Glu | Arg | Arg 270 | His | Lys | |
| 20 | Asp | Ala | Ile 275 | Ala | Ile | Lys | Tyr | Phe 280 | Arg | Lys | Phe | Val | Ala 285 | Ser | Lys | Met | |
| 25 | Ala | Glu 290 | Leu | Ser | Val | Pro | Leu 295 | Asp | Ile | Ile | Asp | Phe 300 | Ile | Gln | Gly | Årg | |
| | Lys 305 | Pro | Thr | Arg | Val | Leu 310 | Thr | Gln | His | Tyr | Val 315 | Ser | Leu | Phe | Gly | Ile 320 | |
| 30 | Ala | Lys | Glu | Gln | Tyr 325 | Lys | Lys | Tyr | Ala | Glu 330 | Trp | Leu | Lys | Gly | Val 335 | | |
| 35 | <211 <212 | 0> 66 L> 14 2> DI 3> A1 | 141 | icia: | L Sec | quenc | ce | | | | | | | | | | |
| 40 | <220 <223 | 3> De | escri oding | iptio | on of | f Art | ific prot | cial tein | Sequ | ience -Xi <i>sl</i> | e: Dì | NA se | equer | ice | | | |
| 45 | | L> CI | os L) | (1437 | 7) | | | | | | | | | | | | |
| 50 | atg |)> 66 ccc Pro | aag Lys | aag Lys | aag Lys 5 | agg Arg | aag Lys | gtg Val | caa Gln | aat Asn 10 | cag Gln | ggt Gly | caa Gln | gac Asp | aaa Lys 15 | tat Tyr | 48 |
| 55 | caa Gln | caa Gln | gcc Ala | ttt Phe 20 | gca Ala | gac Asp | tta Leu | gag Glu | cca Pro 25 | ctt Leu | tca Ser | tct Ser | acc Thr | gac Asp 30 | ggc Gly | agt Ser | 96 |
| ,, | ttt Phe | ctc Leu | ggc Gly 35 | tca Ser | agt Ser | ctg Leu | caa Gln | gca Ala 40 | cag Gln | cag Gln | caa Gln | aga Arg | gaa Glu 45 | cac His | atg Met | aga Arg | 144 |
| 50 | aca Thr | aaa Lys 50 | gta Val | cta Leu | caa Gln | gac Asp | cta Leu 55 | gac Asp | aag Lys | gta Val | aat Asn | ctg Leu 60 | cgt Arg | ttg Leu | aag Lys | tct Ser | 192 |
| 55 | gca Ala 65 | aag Lys | acg Thr | aaa Lys | gtc Val | tca Ser 70 | gtt Val | cga Arg | gaa Glu | tct Ser | aac Asn 75 | gga Gly | agt Ser | ctg Leu | caa Gln | tta Leu 80 | 240 |

| | | | | _ | | | | | | 50 | | | | | | PC1/ | EP01/129 |
|----|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|----------|
| | cga Arg | gca Ala | acg Thr | tta Leu | cca Pro 85 | тте | aaa Lys | cct | gga Gly | 52 gat Asp 90 | aag Lys | gac Asp | acc Thr | aac Asn | ggt Gly 95 | aca Thr | 288 |
| 5 | Gly Ggc | aga Arg | aag Lys | Gln 100 | TĀT | aat Asn | cto Leu | ago Ser | ttg Leu 105 | Asn | ato Ile | cct Pro | gca Ala | aac Asn 110 | Leu | gat Asp | 336 |
| 10 | Gry | nec | 115 | IIII | Ala | GIU | GIU | 120 | Aļa | Tyr | Glu | Leu | 125 | Lys | Leu | atc Ile | 384 |
| 15 | vra | 130 |) Tys | rnr | Pue | GLU | 135 | Asn | . Asp | Lys | Tyr | Leu 140 | Gly | Lys | Gĺu | | 432 |
| 20 | 145 | пλя | пуѕ | Asp | ser | 150 | Thr | 116 | GLY | Asp | Leu 155 | Leu | Glu | Lys | Phe | gca Ala 160 | 480 |
| | GIU | GIU | . Iyr | ttt Phe | ьуs 165 | Thr | His | Lys | Arg | Thr 170 | Thr | Lys | Ser | Glu | His 175 | Thr | 528 |
| 25 | ttt Phe | ttt Phe | tac Tyr | tat Tyr 180 | ttt Phe | tcc Ser | cgc | acc Thr | caa Gln 185 | cga Arg | tat Tyr | acc Thr | aat Asn | tcc Ser 190 | aaa Lys | gat Asp | 576 |
| 30 | tta Leu | gca Ala | acg Thr 195 | gcg Ala | gaa Glu | aat Asn | ctc Leu | atc Ile 200 | aat Asn | tca Ser | att Ile | gag Glu | caa Gln 205 | atc | gat Asp | aaa Lys | 624 |
| 35 | gaa Glu | tgg Trp 210 | gcg Ala | aga Arg | tat Tyr | aat Asn | gcc Ala 215 | gcc Ala | aga Arg | gcc Ala | ata Ile | tca Ser 220 | gct Ala | ttt Phe | tgc Cys | ata Ile | 672 |
| 40 | aca Thr 225 | ttc Phe | aat Asn | ata Ile | gaa Glu | att Ile 230 | gat Asp | ttg Leu | tcc Ser | cag Gln | tat Tyr 235 | tcc Ser | aaa Lys | atg Met | cct Pro | gat Asp 240 | 720 |
| | cgc Arg | aat Asn | tcg Ser | cgc Arg | aac Asn 245 | atc Ile | ccc Pro | aca Thr | gat Asp | gca Ala 250 | gaa Glu | ata Ile | cta Leu | tca Ser | gga Gly 255 | att Ile | 768 |
| 45 | acc Thr | aaa Lys | ttt Phe | gaa Glu 260 | gac Asp | tat Tyr | cța Leu | gtt Val | acc Thr 265 | aga Arg | gga Gly | aat Asn | caa Gln | gtt Val 270 | aat Asn | gaa Glu | 816 |
| 50 | gat Asp | gta Val | aaa Lys 275 | gat Asp | agc Ser | tgg Trp | caa Gln | ctt Leu 280 | tgg Trp | cgc Arg | tgg Trp | aca Thr | tat Tyr 285 | gga Gly | atg Met | tta Leu | 864 |
| 55 | 77.0 | 290 | rite | ggt Gly | ьeu | Arg | 295 | Arg | Glu | Ile | Phe | Ile 300 | Asn | Pro | Asn | Ile | 912 |
| 60 | gat Asp 305 | tgg Trp | tgg Trp | tta Leu | agc Ser | aaa Lys 310 | gag Glu | aat Asn | ata Ile | gac Asp | ctc Leu 315 | aca Thr | tgg Trp | aaa Lys | gta Val | gac Asp 320 | 960 |
| | aaa Lys | gaa Glu | tgt Cys | aaa Lys | act Thr 325 | ggt Gly | gaa Glu | aga Arg | caa Gln | gca Ala 330 | tta Leu | ccc Pro | tta Leu | cat His | aaa Lys 335 | gaa Glu | 1008 |
| 65 | tgg Trp | att Ile | gat Asp | gag Glu 340 | ttt Phe | gat Asp | tta Leu | aga Arg | aat Asn 345 | ccg Pro | aaa Lys | tat Tyr | tta Leu | gaa Glu 350 | atg Met | ctg Leu | 1056 |
| | | | | | | | | | | | | | | | | | |

| 5 | gca Ala | aca Thr | gca Ala 355 | att Ile | agt Ser | aaa Lys | aaa Lys | gat Asp 360 | aaa Lys | aca Thr | aat Asn | cat His | gct Ala 365 | gaa Glu | ata Ile | aca Thr | 1104 |
|----------------|---|--|---|--------------------------------|--------------------------------------|------------------------------------|----------------------------|----------------------------|------------------------------------|--|---------------------------------------|--------------------------------|--------------------------------|--|--------------------------------|------------------------|------|
| | gcc Ala | tta Leu 370 | act Thr | cag Gln | cgt Arg | att Ile | agt Ser 375 | tgg Trp | tgg Trp | ttt Phe | cgg Arg | aaa Lys 380 | gtc Val | gaa Glu | tta Leu | gat Asp | 1152 |
| 10 | ttt Phe 385 | aaa Lys | ccc Pro | tat Tyr | gat Asp | tta Leu 390 | cgt Arg | cac His | gcc Ala | tgg Trp | gca Ala 395 | atc Ile | aga Arg | gcg Ala | cat His | att Ile 400 | 1200 |
| 15 | tta Leu | ggc Gly | ata Ile | cca Pro | atc Ile 405 | aaa Lys | gcg Ala | gcg Ala | gct Ala | gat Asp 410 | aat Asn | ttg Leu | Gl y ggg | cat His | agt Ser 415 | atg Met | 1248 |
| 20 | cag Gln | gtt Val | cat His | aca Thr 420 | caa Gln | acc Thr | tat Tyr | cag Gln | cgc Arg 425 | tgg Trp | ttc Phe | tcg Ser | cta Leu | gat Asp 430 | atg Met | cgg Arg | 1296 |
| 25 | aag Lys | tta Leu | gcg Ala 435 | att Ile | aat Asn | cag Gln | gct Ala | ttg Leu 440 | act Thr | aag Lys | agg Arg | aat Asn | gaa Glu 445 | ttt Phe | gag Glu | gtg Val | 1344 |
| | att Ile | agg Arg 450 | gag Glu | gag Glu | aat Asn | gct Ala | aaa Lys 455 | ttg Leu | cag Gln | ata Ile | gaa Glu | aat Asn 460 | gaa Glu | agg Arg | ttg Leu | agg Arg | 1392 |
| 30 | atg Met 465 | gaa Glu | att Ile | gag Glu | aag Lys | tta Leu 470 | aag Lys | atg Met | gaa Glu | ata Ile | gct Ala 475 | tat Tyr | aag Lys | aat Asn | agt Ser | tg <u>a</u> g | 1441 |
| | | | | | | | | | | | | | | | | | |
| 35 | <211 <212 <213 | | 79 RT ctifi | | L Sec | | | | | | | | | | | | |
| 35 40 | <211 <212 <213 | l> 47 2> PE 3> An 3> De | 79 RT rtifi escri | iptic | | Art | cific | | | | | NA se | equei | nce | | | |
| 40 | <211 <212 <213 <223 | l> 47 2> PB 3> An 3> De CO | 79 RT rtifi escri oding | iption | on of | Art sion | ific prot | tein | NLS- | -Xis <i>l</i> | A | | equer Gln | | Lys 15 | Tyr | |
| | <211 <212 <213 <223 <400 Met | l> 47 2> PP 3> An 3> De co 0> 67 Pro | 79 RT rtifi escri oding 7 Lys | iptio g for Lys | on of fus Lys 5 | Art sion Arg | ific prof | val | NLS- | -XisA Asn 10 | A Gln | Gly | | Asp | 15 | _ | |
| 40 | <211 <212 <213 <223 <400 Met 1 Gln | l> 47 2> PE 3> An 3> De CO D> 67 Pro | 79 RT ctiffiescripding Dding Lys | iption for Lys Phe 20 | Lys 5 | Art sion Arg Asp | Lys | Val Glu | Gln Pro 25 | Asn 10 Leu | Gln Ser | Gly Ser | Gln | Asp Asp 30 | 15 Gly | Ser | |
| 40 | <211 <212 <213 <223 <400 Met 1 Gln | l> 47 2> PF 3> An 3> De co Pro Gln | 79 RT rtiff escri oding 7 Lys Ala Gly 35 | Lys Phe 20 Ser | Lys 5 Ala | Art sion Arg Asp | Lys Leu Gln | Val Glu Ala 40 | Gln Pro 25 Gln | Asn 10 Leu Gln | Gln Ser Gln | Gly Ser Arg | Gln Thr | Asp Asp 30 His | 15 Gly Met | Ser | |
| 40 | <211 <212 <213 <223 <400 Met 1 Gln Phe Thr | l> 47 2> PP 3> An 3> De Co O> 67 Pro Gln Leu Lys 50 | 79 RT rtiff escri oding 7 Lys Ala Gly 35 | Lys Phe 20 Ser | Lys 5 Ala Ser | Artsion Arg Asp Leu Asp | Lys Leu Gln Leu 55 | Val Glu Ala 40 Asp | Gln Pro 25 Gln Lys | Asn 10 Leu Gln Val | Gln Ser Gln Asn | Gly Ser Arg Leu 60 | Gln Thr Glu 45 Arg | Asp 30 His | 15 Gly Met Lys | Ser | |
| 40 45 50 | <211 <212 <213 <223 <400 Met 1 Gln Phe Thr Ala 65 | l> 47 2> PP 3> An 3> De Co O> 67 Pro Gln Leu Lys 50 Lys | 79 RT rtiff escri oding 7 Lys Ala Gly 35 Val | Lys Phe 20 Ser Leu Lys | Lys 5 Ala Ser Gln | Artsion Arg Asp Leu Asp Ser 70 | Lys Leu Gln Leu 55 Val | Val Glu Ala 40 Asp | Gln Pro 25 Gln Lys Glu | Asn 10 Leu Gln Val | Gln Ser Gln Asn Asn 75 | Gly Ser Arg Leu 60 | Gln Thr Glu 45 Arg | Asp 30 His Leu | 15 Gly Met Lys Gln | Ser Arg Ser Leu 80 | |
| 40 45 | <211 <212 <213 <223 <400 Met 1 Gln Phe Thr Ala 65 Arg | L> 47 2> PP 3> An 3> De Co O> 67 Pro Gln Leu Lys 50 Lys | 79 RT rtiff escript oding 7 Lys Ala Gly 35 Val Thr | Lys Phe 20 Ser Leu Lys | Lys 5 Ala Ser Gln Val | Artsion Arg Asp Leu Asp Ser 70 | Lys Leu Gln Leu 55 Val | Val Glu Ala 40 Asp Arg | Gln Pro 25 Gln Lys Glu Gly | Asn 10 Leu Gln Val Ser Asp | Gln Ser Gln Asn Asn 75 | Gly Ser Arg Leu 60 Gly | Gln Thr Glu 45 Arg | Asp 30 His Leu Leu | Gly Met Lys Gln Gly 95 | Ser Arg Ser Leu 80 | |
| 40 45 50 | <211 <212 <213 <223 <400 Met 1 Gln Phe Thr Ala 65 Arg Gly | L> 47 2> PP 3> An 3> De CO O> 67 Pro Gln Leu Lys 50 Lys Ala Arg | 79 RT rtiff escription 7 Lys Ala Gly 35 Val Thr Lys | Lys Phe 20 Ser Leu Lys Leu Gln | Lys Lys Ala Ser Gln Val Pro 85 | Artsion Arg Asp Leu Asp Ser 70 Ile | Lys Leu Gln Leu 55 Val Lys | Val Glu Ala 40 Asp Arg Pro | Gln Pro 25 Gln Lys Glu Gly Leu 105 | Asn 10 Leu Gln Val Ser Asp 90 | Gln Ser Gln Asn Asn T5 Lys | Gly Ser Arg Leu 60 Gly Asp | Gln Thr Glu 45 Arg Ser | Asp 30 His Leu Leu Asn Asn | Gly Met Lys Gln Gly 95 Leu | Ser Arg Ser Leu 80 Thr | |

130 135 54 140

| | | | • | | | | 10. | , | | | | 140 | | | | |
|----|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| 5 | Th: | r Lys | s Lys | s Asp | Ser | Gln 150 | Thr | : Ile | Gly | Asp | Leu 155 | Leu | Glu | Lys | Phe | Ala 160 |
| | | | | | . 100 | , | | | | 170 | • | | | | 175 | |
| 10 | | | | 100 | , | | | | 182 | | | | | 190 | | Asp |
| 15 | | | 190 | , | | | | 200 | | | | | 205 | | | Lys |
| 13 | GIT | 210 | Ala | Arg | Tyr | Asn | Ala 215 | Ala | Arg | Ala | Ile | Ser 220 | Ala | Phe | Cys | Įle |
| 20 | Thr 225 | Phe | Asn | Ile | Glu | 11e 230 | Asp | Leu | Ser | Gln | Tyr 235 | Ser | Lys | Met | Pro | Asp 240 |
| | | | | | Asn 245 | | | | | 250 | | | | | 255 | |
| 25 | | | | 200 | | | | | 265 | | | | | 270 | | |
| | Asp | Val | Lys 275 | Asp | Ser | Trp | Gln | Leu 280 | Trp | Arg | Trp | Thr | Tyr 285 | Gly | Met | Leu |
| 30 | Ala | Val 290 | Phe | Gly | Leu | Arg | Pro 295 | Arg | Glu | Ile | Phe | Ile 300 | Asn | Pro | Asn | Ile |
| 35 | 505 | | | | Ser | 210 | | | | | 315 | • | | | | 320 |
| | | | | • | Thr 325 | | | | | 330 | | | | | 335 | |
| 40 | | | | 340 | | | | | 345 | | | | | 350 | | |
| 45 | | | 333 | | Ser | | | 360 | | | | | 365 | | | |
| 45 | | 370 | | | Arg | | 3/5 | | | | | 380 | | | | |
| 50 | Phe 385 | Lys | Pro | Tyr | Asp | Leu 390 | Arg | His | Ala | Trp | Ala 395 | Ile | Arg | Ala | His | Ile 400 |
| | Leu | Gly | Ile | Pro | Ile 405 | Lys | Ala | Ala | Ala | Asp 410 | Asn | Leu | Gly | His | Ser 415 | Met |
| 55 | Gln | Val | His | Thr 420 | Gln | Thr | Tyr | Gln | Arg 425 | Trp | Phe | Ser | Leu | Asp 430 | Met | Arg |
| | Lys | Leu | Ala 435 | Ile | Asn | Gln | Ala | Leu 440 | Thr | Lys | Arg | Asn | Glu 445 | Phe | Glu | Val |
| 60 | Ile | Arg 450 | Glu | Glu | Asn | Ala | Lys 455 | Leu | Gln | Ile | Glu | Asn 460 | Glu | Arg | Leu | Arg |
| 65 | Met 465 | Glu | Ile | Glu | Lys | Leu 470 | Lys | Met | Glu | Ile | Ala 475 | Tyr | Lys | Asn | Ser | |

a supplied to

<210> 68 <211> 1029

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: DNA sequence coding for fusion protein NLS-Ssv

10 <220>

<221> CDS

<222> (1)..(1026)

<400> 68

15 atg ccc aag aag aag agg aag gtg acg aaa gat aag acg cgt tat aaa 48 Met Pro Lys Lys Lys Arg Lys Val Thr Lys Asp Lys Thr Arg Tyr Lys

tac ggg gat tat att tta cgc gag agg aaa ggg cgg tat tat gtt tac 96 20 Tyr Gly Asp Tyr Ile Leu Arg Glu Arg Lys Gly Arg Tyr Tyr Val Tyr

aag cta gag tat gaa aac ggt gag gta aaa gag cgt tac gtg ggt cct 144 Lys Leu Glu Tyr Glu Asn Gly Glu Val Lys Glu Arg Tyr Val Gly Pro 25

tta gct gac gtc gtt gaa tca tat cta aaa atg aaa tta ggg gtc gta Leu Ala Asp Val Val Glu Ser Tyr Leu Lys Met Lys Leu Gly Val Val 55 30

ggg gat act ecc eta caa geg gat ecc ecc ggt tte gag ecc ggg aca Gly Asp Thr Pro Leu Gln Ala Asp Pro Pro Gly Phe Glu Pro Gly Thr

35 agc gga agc ggt ggt gga aaa gag gga act gaa cga cgt aaa ata gcg 288 Ser Gly Ser Gly Gly Lys Glu Gly Thr Glu Arg Arg Lys Ile Ala

ttg gtt gcc aat ttg cgc caa tac gcg acg gac ggc aac ata aag gcg 40 Leu Val Ala Asn Leu Arg Gln Tyr Ala Thr Asp Gly Asn Ile Lys Ala 105

ttc tac aac tat ctc atg aac gaa agg ggg ata agc gaa aaa act gca 384 Phe Tyr Asn Tyr Leu Met Asn Glu Arg Gly Ile Ser Glu Lys Thr Ala 45 125

aag gac tac atc aat gct ata tca aag ccg tat aaa gag acg aga gac 432 Lys Asp Tyr Ile Asn Ala Ile Ser Lys Pro Tyr Lys Glu Thr Arg Asp 135 50

gca cag aag gct tac cga ctc ttt gca cgt ttc tta gcg tca cgc aat 480 Ala Gln Lys Ala Tyr Arg Leu Phe Ala Arg Phe Leu Ala Ser Arg Asn 155

atc ata cat gat gaa ttt gcg gat aaa ata ttg aaa gcg gta aag gtg 528 Ile Ile His Asp Glu Phe Ala Asp Lys Ile Leu Lys Ala Val Lys Val 165 170

aag aag geg aac get gat ate tae att eea aeg ttg gaa gag ata aaa 576 60 Lys Lys Ala Asn Ala Asp Ile Tyr Ile Pro Thr Leu Glu Glu Ile Lys 180

agg acg tta caa tta gca aaa gac tat agc gaa aac gtc tac ttc atc Arg Thr Leu Gln Leu Ala Lys Asp Tyr Ser Glu Asn Val Tyr Phe Ile 65 195

tac cgt atc gct ctc gag tcg ggc gtt agg ctg agc gaa ata ctg aaa 672 ·

56 Tyr Arg Ile Ala Leu Glu Ser Gly Val Arg Leu Ser Glu Ile Leu Lys gtg ctg aag gaa ccc gaa agg gac att tgc ggt aac gac gtc tgt tat 720 Val Leu Lys Glu Pro Glu Arg Asp Ile Cys Gly Asn Asp Val Cys Tyr 230 tat ccg ctt agt tgg act agg gga tat aag ggc gtc ttc tat gta ttc 768 Tyr Pro Leu Ser Trp Thr Arg Gly Tyr Lys Gly Val Phe Tyr Val Phe 10 cac ata acg cct ctg aag aga gta gag gtg acg aag tgg gca ata gcg 816 His Ile Thr Pro Leu Lys Arg Val Glu Val Thr Lys Trp Ala Ile Ala 260 265 15 gac ttt gaa cga cgt cat aag gac gct ata gcg ata aag tac ttc cgc 864 Asp Phe Glu Arg Arg His Lys Asp Ala Ile Ala Ile Lys Tyr Phe Arg 280 285 20 aaa ttc gta gcg tct aag atg gct gag cta agc gta ccg tta gat att 912 Lys Phe Val Ala Ser Lys Met Ala Glu Leu Ser Val Pro Leu Asp Ile 290 . 295 atc gat ttt att caa ggg cgt aaa ccg aca cgc gtt tta acg caa cat Ile Asp Phe Ile Gln Gly Arg Lys Pro Thr Arg Val Leu Thr Gln His 310 tac gta tcg ctc ttc ggc ata gcg aaa gag caa tat aaa aag tat gcg 1008 Tyr Val Ser Leu Phe Gly Ile Ala Lys Glu Gln Tyr Lys Lys Tyr Ala 30 325 gaa tgg cta aaa ggg gtc tga 1029 Glu Trp Leu Lys Gly Val 340 35 <210> 69 <211> 342 <212> PRT <213> Artificial Sequence <223> Description of Artificial Sequence: DNA sequence coding for fusion protein NLS-Ssv <400> 69 45 Met Pro Lys Lys Lys Arg Lys Val Thr Lys Asp Lys Thr Arg Tyr Lys 10 Tyr Gly Asp Tyr Ile Leu Arg Glu Arg Lys Gly Arg Tyr Tyr Val Tyr 50 Lys Leu Glu Tyr Glu Asn Gly Glu Val Lys Glu Arg Tyr Val Gly Pro Leu Ala Asp Val Val Glu Ser Tyr Leu Lys Met Lys Leu Gly Val Val 55 Gly Asp Thr Pro Leu Gln Ala Asp Pro Pro Gly Phe Glu Pro Gly Thr Ser Gly Ser Gly Gly Gly Lys Glu Gly Thr Glu Arg Arg Lys Ile Ala 90 Leu Val Ala Asn Leu Arg Gln Tyr Ala Thr Asp Gly Asn Ile Lys Ala 105 65 Phe Tyr Asn Tyr Leu Met Asn Glu Arg Gly Ile Ser Glu Lys Thr Ala

125

```
Lys Asp Tyr Ile Asn Ala Ile Ser Lys Pro Tyr Lys Glu Thr Arg Asp
     Ala Gln Lys Ala Tyr Arg Leu Phe Ala Arg Phe Leu Ala Ser Arg Asn
     Ile Ile His Asp Glu Phe Ala Asp Lys Ile Leu Lys Ala Val Lys Val
                                          170
 10
     Lys Lys Ala Asn Ala Asp Ile Tyr Ile Pro Thr Leu Glu Glu Ile Lys
     Arg Thr Leu Gln Leu Ala Lys Asp Tyr Ser Glu Asn Val Tyr Phe Ile
15
     Tyr Arg Ile Ala Leu Glu Ser Gly Val Arg Leu Ser Glu Ile Leu Lys
20
     Val Leu Lys Glu Pro Glu Arg Asp Ile Cys Gly Asn Asp Val Cys Tyr
     Tyr Pro Leu Ser Trp Thr Arg Gly Tyr Lys Gly Val Phe Tyr Val Phe
                     245
25.
     His Ile Thr Pro Leu Lys Arg Val Glu Val Thr Lys Trp Ala Ile Ala
     Asp Phe Glu Arg Arg His Lys Asp Ala Ile Ala Ile Lys Tyr Phe Arg
30
     Lys Phe Val Ala Ser Lys Met Ala Glu Leu Ser Val Pro Leu Asp Ile
35
     Ile Asp Phe Ile Gln Gly Arg Lys Pro Thr Arg Val Leu Thr Gln His
                         310
     Tyr Val Ser Leu Phe Gly Ile Ala Lys Glu Gln Tyr Lys Lys Tyr Ala
                     325
                                                              335
40
     Glu Trp Leu Lys Gly Val
                 340
45
     <210> 70
     <211> 3908
     <212> DNA
     <213> Artificial Sequence
50
     <220>
     <223> Description of Artificial Sequence: vector
           pBS-SSV3
55
     <400> .70
     cacctaaatt gtaagcgtta atattttgtt aaaattcgcg ttaaattttt gttaaatcag 60
     ctcattttt aaccaatagg ccgaaatcgg caaaatccct tataaatcaa aagaatagac 120
     cgagataggg ttgagtgttg ttccagtttg gaacaagagt ccactattaa agaacgtgga 180
     ctccaacgtc aaagggcgaa aaaccgtcta tcagggcgat ggcccactac gtgaaccatc 240
60
    accctaatca agtitttgg ggtcgaggtg ccgtaaagca ctaaatcgga accctaaagg 300
    gagcccccga tttagagctt gacggggaaa gccggcgaac gtggcgagaa aggaagggaa 360
     gaaagcgaaa ggagcgggcg ctagggcgct ggcaagtgta gcggtcacgc tgcgcgtaac 420
    caccacacce geogegetta atgegeeget acagggegeg teccattege catteagget 480
     gegeaactgt tgggaaggge gateggtgeg ggeetetteg etattaegee agetggegaa 540
     agggggatgt getgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg 600
    ttgtaaaacg acggccagtg aattgtaata cgactcacta tagggcgaat tggagctcca 660
    ccgcggtggc ggccgcccga tatgacgaaa gataagacgc gttataaata cggggattat 720
```

```
attttacgcg agaggaaagg gcggtattat gtttacaagc tagagtatga aaacggtgag 780
      gtaaaagagc gttacgtggg tcctttagct gacgtcgttg aatcatatct aaaaatgaaa 840
      ttaggggtcg taggggatac tcccctacaa gcggatcccc ccggtttcga gcccgggaca 900
      agcggaagcg gtggtggaaa agagggaact gaacgacgta aaatagcgtt ggttgccaat 960
      ttgcgccaat acgcgacgga cggcaacata aaggcgttct acaactatct catgaacgaa 1020
      agggggataa gcgaaaaac tgcaaaggac tacatcaatg ctatatcaaa gccgtataaa 1080
      gagacgagag acgcacagaa ggcttaccga ctctttgcac gtttcttagc gtcacgcaat 1140
      atcatacatg atgaattigc ggataaaata ttgaaagcgg taaaggtgaa gaaggcgaac 1200
      gctgatatct acattccaac gttggaagag ataaaaagga cgttacaatt agcaaaagac 1260
 10
      tatagegaaa aegtetaett catetaeegt ategeteteg agtegggegt taggetgage 1320
      gaaatactga aagtgctgaa ggaacccgaa agggacattt gcggtaacga cgtctgttat 1380
      tatccgctta gttggactag gggatataag ggcgtcttct atgtattcca cataacgcct 1440
      ctgaagagag tagaggtgac gaagtgggca atagcggact ttgaacgacg tcataaggac 1500
      getatagega taaagtaett eegeaaatte gtagegteta agatggetga getaagegta 1560
 15
      ccgttagata ttatcgattt tattcaaggg cgtaaaccga cacgcgtttt aacgcaacat 1620
      tacgtatcgc tcttcggcat agcgaaagag caatataaaa agtatgcgga atggctaaaa 1680
     ggggtctgac tcgagggggg gcccggtacc cagcttttgt tccctttagt gagggttaat 1740
      ttogagettg gegtaateat ggteataget gttteetgtg tgaaattgtt atcegeteae 1800
     aattccacac aacatacgag ccggaagcat aaagtgtaaa gcctggggtg cctaatgagt 1860
20
     gagctaactc acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc 1920
     gtgccagctg cattaatgaa tcggccaacg cgcgggaga ggcggtttgc gtattgggcg 1980
     ctetteeget teetegetea etgaeteget gegeteggte gtteggetge ggegageggt 2040
     atcageteae teaaaggegg taataeggtt atceaeagaa teaggggata acgeaggaaa 2100
     gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc 2160
25
     gtttttccat aggeteegee eeeetgaega geateacaaa aategaeget caagteagag 2220
     gtggcgaaac ccgacaggac tataaagata ccaggcgttt ccccctggaa gctccctcgt 2280
     gegeteteet gtteegaeee tgeegettae eggataeetg teegeettte teeetteggg 2340
     aagegtggeg ettteteata geteaegetg taggtatete agtteggtgt aggtegtteg 2400
     ctccaagetg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccgg 2460
30
     taactategt ettgagteea acceggtaag acacgaetta tegecaetgg cageageeac 2520
     tggtaacagg attagcagag cgaggtatgt aggcggtgct acagagttct tgaagtggtg 2580.
     geetaactae ggetacaeta gaaggacagt atttggtate tgegetetge tgaageeagt 2640
     taccttcgga aaaagagttg gtagetcttg atccggcaaa caaaccaccg ctggtagcgg 2700
     tggtttttt gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc 2760
35
     titgatettt tetacggggt etgacgetea gtggaacgaa aacteacgtt aagggatttt 2820
     taaatcaatc taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag 2940
     tgaggeacet ateteagega tetgtetatt tegtteatee atagttgeet gaeteeeegt 3000
     cgtgtagata actacgatac gggagggett accatetgge eccagtgetg caatgatace 3060
40
     gcgagaccca cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc 3120
     cgagcgcaga agtggtcctg caactttatc cgcctccatc cagtctatta attgttgccg 3180
     ggaagctaga gtaagtagtt cgccagttaa tagtttgcgc aacgttgttg ccattgctac 3240
     aggeategig gigleacget egiegitigg talggetica ticageticg gitteceaacg 3300
     atcaaggcga gttacatgat cocceatgtt gtgcaaaaaa gcggttagct ccttcggtcc 3360
45
     tccgatcgtt gtcagaagta agttggccgc agtgttatca ctcatggtta tggcagcact 3420
     gcataattet ettaetgtea tgecateegt aagatgettt tetgtgaetg gtgagtaete 3480
     aaccaagtca ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat 3540 acgggataat accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc 3600
     ttcggggcga aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaacccac 3660
50
     togtgcacco aactgatott cagcatottt tactttcacc agogtttctg ggtgagcaaa 3720
     aacaggaagg caaaatgccg caaaaaaggg aataagggcg acacggaaat gttgaatact 3780
     catactette ettttteaat attattgaag catttateag ggttattgte teatgagegg 3840
     atacatattt gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg 3900
     aaaagtgc
55
     <210> 71
     <211> 3927
     <212> DNA
60
     <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: vector
          pBS-SSV4
65
     <400> 71
     cacctaaatt gtaagcgtta atattttgtt aaaattcgcg ttaaattttt gttaaatcag 60
```

WO 02/38613

59

PCT/EP01/12975

ctcatttttt aaccaatagg ccgaaatcgg caaaatccct tataaatcaa aagaatagac 120 cgagataggg ttgagtgttg ttccagtttg gaacaagagt ccactattaa agaacgtgga 180 ctccaacgtc aaagggcgaa aaaccgtcta tcagggcgat ggcccactac gtgaaccatc 240 accetaatea agttttttgg ggtcgaggtg ccgtaaagca ctaaategga accetaaagg 300 gagcccccga tttagagctt gacggggaaa gccggcgaac gtggcgagaa aggaagggaa 360 gaaagegaaa ggagegggeg etagggeget ggeaagtgta geggteaege tgegegtaac 420 caecacace geegegetta atgegeeget acagggegeg teccattege catteagget 480 gcgcaactgt tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa 540 agggggatgt gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg 600 ttgtaaaacg acggccagtg aattgtaata cgactcacta tagggcgaat tggagctcca 660 ccgcggtggc ggccgcacca tgcccaagaa gaagaggaag gtgacgaaag ataagacgcg 720 ttataaatac ggggattata ttttacgcga gaggaaaggg cggtattatg tttacaagct 780 agagtatgaa aacggtgagg taaaagagcg ttacgtgggt cctttagctg acgtcgttga 840 atcatatcta aaaatgaaat taggggtcgt aggggatact cccctacaag cggatccccc 900 15 cggtttcgag cccgggacaa gcggaagcgg tggtggaaaa gagggaactg aacgacgtaa 960 aatagegttg gttgecaatt tgegecaata egegaeggae ggeaacataa aggegtteta 1020 caactatctc atgaacgaaa gggggataag cgaaaaaact gcaaaggact acatcaatgc 1080 tatatcaaag ccgtataaag agacgagaga cgcacagaag gcttaccgac tctttgcacg 1140 tttcttagcg tcacgcaata tcatacatga tgaatttgcg gataaaatat tgaaagcggt 1200 20 aaaggtgaag aaggcgaacg ctgatatcta cattccaacg ttggaagaga taaaaaggac 1260 gttacaatta gcaaaagact atagcgaaaa cgtctacttc atctaccgta tcgctctcga 1320 gtcgggcgtt aggctgagcg aaatactgaa agtgctgaag gaacccgaaa gggacatttg 1380 cggtaacgac gtctgttatt atccgcttag ttggactagg ggatataagg gcgtcttcta 1440 tgtattccac ataacgcctc tgaagagagt agaggtgacg aagtgggcaa tagcggactt 1500 25 tgaacgacgt cataaggacg ctatagcgat aaagtacttc cgcaaattcg tagcgtctaa 1560 gatggctgag ctaagcgtac cgttagatat tatcgatttt attcaagggc gtaaaccgac 1620 acgcgtttta acgcaacatt acgtatcgct cttcggcata gcgaaagagc aatataaaaa 1680 gtatgeggaa tggetaaaag gggtetgaet egaggggggg ceeggtaeee agettttgtt 1740 eeetttagtg agggttaatt tegagettgg egtaateatg gteatagetg ttteetgtgt 1800 30 gaaattgtta teegeteaca atteeacaca acatacgage eggaageata aagtgtaaag 1860 cctggggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt 1920 tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag 1980 geggtttgeg tattgggege tetteegett cetegeteae tgactegetg egeteggteg 2040 ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta tccacagaat 2100 35 caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 2160 aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa 2220 atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 2280 cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt 2340 ccgcctttet cccttcggga agcgtggcgc tttctcatag ctcacgctgt aggtatctca 2400 40 gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaacccccc gttcagcccg 2460 accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat 2520 cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 2580 cagagttett gaagtggtgg cetaactacg getacactag aaggacagta tttggtatet 2640 gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac 2700 45 aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa 2760 aaggatetea agaagateet ttgatetttt etaeggggte tgaegeteag tggaacgaaa 2820 actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt 2880 taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tggtctgaca 2940 gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca 3000 50 tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc 3060 ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta tcagcaataa 3120 accagecage eggaagggee gagegeagaa gtggteetge aactttatee geetecatee 3180 agtetattaa ttgttgccgg gaagetagag taagtagtte geeagttaat agtttgcgca 3240 acgttgttgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt atggcttcat 3300 55 teageteegg tteceaacga teaaggegag ttacatgate ecceatgttg tgeaaaaaag 3360 cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac 3420 tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt 3480 ctgtgactgg tgagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt 3540 getettgece ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc 3600 60 teateattgg aaaacgttet teggggegaa aacteteaag gatettaceg etgttgagat 3660 ccagttegat gtaacccact cgtgcaccca actgatette agcatetttt actttcacca 3720 gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga 3780 cacggaaatg ttgaatactc atactettee tttttcaata ttattgaage atttatcagg 3840 gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa caaatagggg 3900 65 ttccgcgcac atttccccga aaagtgc 3927

```
<210> 72
     <211> 3351
     <212> DNA
     <213> Artificial Sequence
 5
     <220>
     <223> Description of Artificial Sequence: vector
           pBS-SSVs
10
     <400> 72
     cgacctcgag tcagacccct tttagccatt ccgcatactt tttatattgc tctttcgcta 60
     tgccgaagag cgatacgtaa tgttgcgtta aaacgcgtgt cggtttacgc ccttgaataa 120
     aatcgataat atctaacggt acgcttagct cagccatctt agacgctacg aatttgcgga 180
     agtactttat cgctatagcg tccttatgac gtcgttcaaa gtccgctatt gcccacttcg 240
15
     teacetetae tetetteaga ggegttatgt ggaatacata gaagacgeee ttatateeee 300
     tagtecaact aageggataa taacagacgt cgttaccgca aatgtecett tegggtteet 360
     tcagcacttt cagtatttcg ctcagcctaa cgcccgactc gagggggggc ccggtaccca 420
     gcttttgttc cctttagtga gggttaattt cgagcttggc gtaatcatgg tcatagctgt 480
     ttcctgtgtg aaattgttat ccgctcacaa ttccacacaa catacgagec ggaagcataa 540
20
     agtgtaaagc ctggggtgcc taatgagtga gctaactcac attaattgcg ttgcgctcac 600
     tgcccgcttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg 660
     cggggagagg cggtttgcgt attgggcgct cttccgcttc ctcgctcact gactcgctgc 720
     gctcggtcgt tcggctgcgg cgagcggtat cagctcactc aaaggcggta atacggttat 780
     ccacagaatc aggggataac gcaggaaaga acatgtgagc aaaaggccag caaaaggcca 840
25
     ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc cctgacgagc 900
     atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta taaagatacc 960
     aggogtttee ecctggaage tecetegtge geteteetgt tecgaceetg ecgettaceg 1020
     gatacetgte egeetttete eettegggaa gegtggeget tteteatage teaegetgta 1080
     ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg 1140
30
     ttcagcccga ccgctgcgcc ttatccggta actatcgtct tgagtccaac ccggtaagac 1200
     acgaettate gecaetggea geagecaetg gtaacaggat tageagageg aggtatgtag 1260
     gcggtgctac agagttcttg aagtggtggc ctaactacgg ctacactaga aggacagtat 1320
     ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt agctcttgat 1380
     ccggcaaaca aaccaccgct ggtagcggtg gttttttgt ttgcaagcag cagattacgc 1440 gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacggggtct gacgctcagt 1500
35
     ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt atcaaaaagg atcttcacct 1560
     agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat gagtaaactt 1620
     ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtctatttc 1680
     gttcatccat agttgcctga ctccccgtcg tgtagataac tacgatacgg gagggcttac 1740
40
     catctggccc cagtgctgca atgataccgc gagacccacg ctcaccggct ccagatttat 1800
     cagcaataaa ccagccagcc ggaagggccg agcgcagaag tggtcctgca actttatccg 1860
     cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttaata 1920
     gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt gtcacgctcg tcgtttggta 1980
     tggcttcatt cagctccggt tcccaacgat caaggcgagt tacatgatcc cccatgttgt 2040
45
     gcaaaaaagc ggttagctcc tteggtcctc cgatcgttgt cagaagtaag ttggccgcag 2100
     tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg ccatccgtaa 2160
     gatgetttte tgtgactggt gagtactcaa ccaagtcatt ctgagaatag tgtatgegge 2220
     gaccgagttg ctcttgcccg gcgtcaatac gggataatac cgcgccacat agcagaactt 2280
     taaaagtgct catcattgga aaacgttett eggggegaaa acteteaagg atettaeege 2340
    tgttgagatc cagttcgatg taacccactc gtgcacccaa ctgatcttca gcatctttta 2400
    ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa 2460
     taagggcgac acggaaatgt tgaatactca tactcttcct ttttcaatat tattgaagca 2520
     tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag aaaaataaac 2580
     aaataggggt teegegeaca ttteeeegaa aagtgeeace taaattgtaa gegttaatat 2640
55
    tttgttaaaa ttcgcgttaa attttgtta aatcagctca ttttttaacc aataggccga 2700
     aatcggcaaa atcccttata aatcaaaaga atagaccgag atagggttga gtgttgttcc 2760
    agtttggaac aagagtccac tattaaagaa cgtggactcc aacgtcaaag ggcgaaaaac 2820
    cgtctatcag ggcgatggcc cactacgtga accatcaccc taatcaagtt ttttggggtc 2880
    gaggtgccgt aaagcactaa atcggaaccc taaagggagc ccccgattta gagcttgacg 2940
60
    gggaaagccg gcgaacgtgg cgagaaagga agggaagaaa gcgaaaggag cgggcgctag 3000
    ggcgctggca agtgtagcgg tcacgctgcg cgtaaccacc acacccgccg cgcttaatgc 3060
     gccgctacag ggcgcgtccc attcgccatt caggctgcgc aactgttggg aagggcgatc 3120
    ggtgcgggcc tcttcgctat tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt 3180
    aagttgggta acgccagggt tttcccagtc acgacgttgt aaaacgacgg ccagtgaatt 3240
65
     gtaatacgac tcactatagg gcgaattgga gctccaccgc ggtggcggcc gctctagaac 3300
    tagtggatcc cccgggctgc aggaattcga tatcaagctt atcgataccg t
```

```
<210> 73
      <211> 5730
      <212> DNA
  5
      <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: vector
            pCMVC31 (NNLS)
 10
      <400> 73
      aaacagtccg atgtacgggc cagatatacg cgttgacatt gattattgac tagttattaa 60
      tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg cgttacataa 120
      cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt gacgtcaata 180
 15
      atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca atgggtggac 240
      tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc aagtacgccc 300
      cotattgacg toaatgacgg taaatggccc gcctggcatt atgcccagta catgacctta 360
      tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac catggtgatg 420
      cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg atttccaagt 480
20
      ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg ggactttcca 540
      aaatgtegta acaacteege eecattgaeg caaatgggeg gtaggegtgt acggtgggag 600
      gtctatataa gcagagctet ctggctaact agagaaceca etgettaetg gettategaa 660
      attaatacga etcaetatag ggagacceaa getgaeteta gaettaatta agegttgggg 720
     tgagtactcc ctctcaaaag cgggcatgac ttctgcgcta agattgtcag tttccaaaaa 780
25
     cgaggaggat ttgatattca cctggcccgc ggtgatgcct ttgagggtgg ccgcgtccat 840
     ctggtcagaa aagacaatct ttttgttgtc aagcttgagg tgtggcaggc ttgagatctg 900
     gccatacact tgagtgacat tgacatccac tttgcctttc tctccacagg tgtccactcc 960
     cagggeggee geaceatgee caagaagaag aggaaggtga cacaaggggt tgtgaeeggg 1020
     gtggacacgt acgcgggtgc ttacgaccgt cagtcgcgcg agcgcgagaa ttcgagcgca 1080
30
     gcaagcccag cgacacagcg tagcgccaac gaagacaagg cggccgacct tcagcgcgaa 1140
     gtcgagcgcg acgggggccg gttcaggttc gtcgggcatt tcagcgaagc gccgggcacg 1200
     teggegtteg ggaeggegga gegeeeggag ttegaacgea teetgaacga atgeeggee 1260 gggeggetea acatgateat tgtetatgae gtgtegeget tetegeget gaaggteatg 1320
     gacgcgattc cgattgtctc ggaattgctc gccctgggcg tgacgattgt ttccactcag 1380
35
     gaaggegtet teeggeaggg aaaegteatg gaeetgatte acetgattat geggetegae 1440
     gegtegeaca aagaatette getgaagteg gegaagatte tegacaegaa gaacetteag 1500
     cgcgaattgg gcgggtacgt cggcgggaag gcgccttacg gcttcgagct tgtttcggag 1560
     acgaaggaga tcacgcgcaa cggccgaatg gtcaatgtcg tcatcaacaa gcttgcgcac 1620
     tegaccacte ceettacegg accettegag ttegageceg acgtaateeg gtggtggtgg 1680
40
     cgtgagatca agacgcacaa acacetteee tteaageegg geagtcaage egecatteae 1740
     cogggcagca tcacggggct ttgtaagcgc atggacgctg acgccgtgcc gacccggggc 1800
     gagacgattg ggaagaagac cgcttcaagc gcctgggacc cggcaaccgt tatgcgaatc 1860
     cttcgggacc cgcgtattgc gggcttcgcc gctgaggtga tctacaagaa gaagccggac 1920
     ggcacgccga ccacgaagat tgagggttac cgcattcagc gcgacccgat cacgctccgg 1980
45
     ccggtcgagc ttgattgcgg accgatcate gagcccgctg agtggtatga gcttcaggcg 2040
     tggttggacg gcagggggcg cggcaagggg ctttcccggg ggcaagccat tctgtccgcc 2100 atggacaagc tgtactgcga gtgtggcgcc gtcatgactt cgaagcgcgg ggaagaatcg 2160 atcaaggact cttaccgctg ccgtcgccgg aaggtggtcg acccgtccgc acctgggcag 2220
     cacgaaggca cgtgcaacgt cagcatggcg gcactcgaca agttcgttgc ggaacgcatc 2280
50
     ttcaacaaga tcaggcacgc cgaaggcgac gaagagacgt tggcgcttct gtgggaagcc 2340
     gcccgacget teggcaaget cactgaggeg cetgagaaga geggegaacg ggcgaacett 2400
     gttgcggagc gcgccgacgc cctgaacgcc cttgaagagc tgtacgaaga ccgcgcggca 2460
     ggcgcgtacg acggacccgt tggcaggaag cacttccgga agcaacaggc agcgctgacg 2520
     ctccggcagc aaggggcgga agagcggctt gccgaacttg aagccgccga agccccgaag 2580
55
     ctteccettg accaatggtt ccccgaagac geegacgctg accegacegg ccctaagteg 2640
     tggtggggge gegegteagt agacgacaag egegtgtteg tegggetett egtagacaag 2700
     atcgttgtca cgaagtcgac tacgggcagg gggcagggaa cgcccatcga gaagcgcgct 2760
     togateacgt gggcgaagec gecgacegae gacgacgaag acgacgecea ggacggcaeg 2820
     gaagacgtag cggcgtaggc ggcgcceggg ctcgagatcc aggcgcggat caataaaaga 2880
     tcattatttt caatagatct gtgtgttggt tttttgtgtg ccttggggga gggggaggcc 2940
     agaatgaggc gcggccaagg gggagggga ggccagaatg accttggggg agggggaggc 3000
     cagaatgacc ttgggggagg gggaggccag aatgaggcgc gccccgggt accgagctcg 3060
     aatteactgg eegtegtttt acaaegtegt gactgggaaa accetggegt tacceaactt 3120
     aategeetig cagcacatee ceettiegee agetggegta atagegaaga ggeeegeace 3180
65
     gategeeett eccaacagtt gegeageetg aatggegaat ggegeetgat geggtatttt 3240
     cteettacge atotgtgcgg tatttcacae egcatatggt geacteteag tacaatetge 3300
     tetgatgeeg catagttaag ceageceega caecegecaa caecegetga egegeeetga 3360
```

```
cgggcttgtc tgctcccggc atccgcttac agacaagctg tgaccgtctc cgggagctgc 3420
     atgtgtcaga ggttttcacc gtcatcaccg aaacgcgcga gacgaaaggg cctcgtgata 3480
     cgcctatttt tataggttaa tgtcatgata ataatggttt cttagacgtc aggtggcact 3540
     tttcggggaa atgtgcgcgg aacccctatt tgtttatttt tctaaataca ttcaaatatg 3600
     tatccgctca tgagacaata accctgataa atgcttcaat aatattgaaa aaggaagagt 3660
     atgagtattc aacatttccg tgtcgccctt attccctttt ttgcggcatt ttgccttcct 3720
     gtttttgctc acccagaaac gctggtgaaa gtaaaagatg ctgaagatca gttgggtgca 3780
     cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag ttttcgcccc 3840
     gaagaacgtt ttccaatgat gagcactttt aaagttctgc tatgtggcgc ggtattatcc 3900
10
     cgtattgacg ccgggcaaga gcaactcggt cgccgcatac actattctca gaatgacttg 3960
     gttgagtact caccagtcac agaaaagcat cttacggatg gcatgacagt aagagaatta 4020
     tgcagtgctg ccataaccat gagtgataac actgcggcca acttacttct gacaacgatc 4080
     ggaggaccga aggagctaac cgcttttttg cacaacatgg gggatcatgt aactcgcctt 4140
     gatcgttggg aaccggagct gaatgaagcc ataccaaacg acgagcgtga caccacgatg 4200
15
     cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg gcgaactact tactctagct 4260 tcccggcaac aattaataga ctggatggag gcggataaag ttgcaggacc acttctgcgc 4320
     teggecette eggetggetg gtttattget gataaatetg gageeggtga gegtgggtet 4380
     cgcggtatca ttgcagcact ggggccagat ggtaagccct cccgtatcgt agttatctac 4440
     acgacgggga gtcaggcaac tatggatgaa cgaaatagac agatcgctga gataggtgcc 4500
20 tcactgatta agcattggta actgtcagac caagtttact catatatact ttagattgat 4560
     ttaaaacttc atttttaatt taaaaggatc taggtgaaga teetttttga taateteatg 4620
     accaaaatcc cttaacgtga gttttcgttc cactgagcgt cagaccccgt agaaaagatc 4680
     aaaggatett ettgagatee ttttttetg egegtaatet getgettgea aacaaaaaa 4740
     ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct ttttccgaag 4800
25
     gtaactggct tcagcagagc gcagatacca aatactgtcc ttctagtgta gccgtagtta 4860
     ggccaccact tcaagaacte tgtagcaccg cctacatace tegetetget aateetgtta 4920
     ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg ggttggactc aagacgatag 4980
     ttaccggata aggcgcagcg gtcgggctga acggggggtt cgtgcacaca gcccagcttg 5040
     gagcgaacga cctacaccga actgagatac ctacagcgtg agctatgaga aagcgccacg 5100
30
     cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg aacaggagag 5160 cgcacgaggg agcttccagg gggaaacgcc tggtatcttt atagtcctgt cgggtttcgc 5220
     cacctctgac ttgagcgtcg atttttgtga tgctcgtcag gggggcggag cctatggaaa 5280
     aacgccagca acgcggcctt tttacggttc ctggcctttt gctggccttt tgctcacatg 5340
     ttettteetg egttateece tgattetgtg gataacegta ttacegeett tgagtgaget 5400
     gataccgctc gccgcagccg aacgaccgag cgcagcgagt cagtgagcga ggaagcggaa 5460
     gagcgcccaa tacgcaaacc gcctctcccc gcgcgttggc cgattcatta atgcagctgg 5520
     cacgacaggt ttcccgactg gaaagcgggc agtgagcgca acgcaattaa tgtgagttag 5580
     ctcactcatt aggcacccca ggctttacac tttatgcttc cggctcgtat gttgtgtgga 5640
     attgtgagcg gataacaatt tcacacagga aacagctatg accatgatta cgccaagcta 5700
40
     gcccgggcta gcttgcatgc ctgcaggttt
     <210> 74
     <211> 4886
45
     <212> DNA
     <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: vector
50
           pCMV-SSV
     <400> 74
     aaacagtccg atgtacgggc cagatatacg cgttgacatt gattattgac tagttattaa 60
     tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg cgttacataa 120
55
     cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt gacgtcaata 180
     atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca atgggtggac 240
     tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc aagtacgccc 300
     cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta catgacctta 360
     tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac catggtgatg 420
60
     cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg atttccaagt 480
     ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg ggactttcca 540
     aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt acggtgggag 600
     gtctatataa gcagagetet etggetaaet agagaaceea etgettaetg gettategaa 660
     attaatacga ctcactatag ggagacccaa gctgactcta gacttaatta agcgttgggg 720
65
     tgagtactcc ctctcaaaag cgggcatgac ttctgcgcta agattgtcag tttccaaaaa 780
     cgaggaggat ttgatattca cctggcccgc ggtgatgcct ttgagggtgg ccgcgtccat 840
     ctggtcagaa aagacaatct ttttgttgtc aagcttgagg tgtggcaggc ttgagatctg 900
```

gccatacact tgagtgacat tgacatccac tttgcctttc tctccacagg tgtccactcc 960 cagggcggcc gcccgatatg acgaaagata agacgcgtta taaatacggg gattatattt 1020 tacgcgagag gaaagggcgg tattatgttt acaagctaga gtatgaaaac ggtgaggtaa 1080 aagagcgtta cgtgggtcct ttagctgacg tcgttgaatc atatctaaaa atgaaattag 1140 gggtcgtagg ggatactccc ctacaagcgg atcccccgg tttcgagccc gggacaagcg 1200 gaagcggtgg tggaaaagag ggaactgaac gacgtaaaat agcgttggtt gccaatttgc 1260 gccaatacgc gacggacggc aacataaagg cgttctacaa ctatctcatg aacgaaaggg 1320 ggataagcga aaaaactgca aaggactaca tcaatgctat atcaaagccg tataaagaga 1380 cgagagacge acagaagget taccgaetet ttgcacgttt ettagegtea egeaatatea 1440 10 tacatgatga atttgcggat aaaatattga aagcggtaaa ggtgaagaag gcgaacgctg 1500 atatctacat tccaacgttg gaagagataa aaaggacgtt acaattagca aaagactata 1560 gcgaaaacgt ctacttcatc taccgtatcg ctctcgagtc gggcgttagg ctgagcgaaa 1620 tactgaaagt gctgaaggaa cccgaaaggg acatttgcgg taacgacgtc tgttattatc 1680 cgcttagttg gactagggga tataagggcg tcttctatgt attccacata acgcctctga 1740 15 agagagtaga ggtgacgaag tgggcaatag cggactttga acgacgtcat aaggacgcta 1800 tagcgataaa gtacttccgc aaattcgtag cgtctaagat ggctgagcta agcgtaccgt 1860 tagatattat cgattttatt caagggcgta aaccgacacg cgttttaacg caacattacg 1920 tatcgctctt cggcatagcg aaagagcaat ataaaaagta tgcggaatgg ctaaaagggg 1980 tetgaetega gggggggeee gtegaeeteg agatecagge geggateaat aaaagateat 2040 20 tattttcaat agatctgtgt gttggttttt tgtgtgcctt gggggagggg gaggccagaa 2100 tgaggcgcgg ccaaggggga gggggaggcc agaatgacct tgggggaggg ggaggccaga 2160 atgacettgg gggaggggga ggccagaatg aggegegeee eegggtaeeg agetegaatt 2220 cactggccgt cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc 2280 gccttgcagc acatccccct ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc 2340 25 gecettecea acagttgege agectgaatg gegaatggeg cetgatgegg tattttetee 2400 ttacgcatct gtgcggtatt tcacaccgca tatggtgcac tctcagtaca atctgctctg 2460 atgeografia gttaagecag ceeegacace egecaacace egetgaegeg eeetgaeggg 2520 cttgtctgct cccggcatcc gcttacagac aagctgtgac cgtctccggg agctgcatgt 2580 gtcagaggtt ttcaccgtca tcaccgaaac gcgcgagacg aaagggcctc gtgatacgcc 2640 30 tattittata ggttaatgtc atgataataa tggtttctta gacgtcaggt ggcacttttc 2700 ggggaaatgt gcgcggaacc cctatttgtt tatttttcta aatacattca aatatgtatc 2760 cgctcatgag acaataaccc tgataaatgc ttcaataata ttgaaaaagg aagagtatga 2820 gtattcaaca tttccgtgtc gcccttattc ccttttttgc ggcattttgc cttcctgttt 2880 ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag 2940 35 tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagtttt cgccccgaag 3000 aacgttttcc aatgatgagc acttttaaag ttctgctatg tggcgcggta ttatcccgta 3060 ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggttg 3120 agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca 3180 gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcggag 3240 40 gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgccttgatc 3300 gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 3360 tagcaatggc aacaacgttg cgcaaactat taactggcga actacttact ctagcttccc 3420 ggcaacaatt aatagactgg atggaggcgg ataaagttgc aggaccactt ctgcgctcgg 3480 cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt gggtctcgcg 3540 45 gtatcattgc agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacga 3600 cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 3660 tgattaagca ttggtaactg tcagaccaag tttactcata tatactttag attgatttaa 3720 aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca 3780 aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag 3840 50 gatettettg agateetttt tttetgegeg taatetgetg ettgeaaaca aaaaaaccac 3900 cgctaccage ggtggtttgt ttgccggate aagagctace aactetttt ccgaaggtaa 3960 ctggcttcag cagagcgcag ataccaaata ctgtccttct agtgtagccg tagttaggcc 4020 accaetteaa gaaetetgta geaeegeeta eataeetege tetgetaate etgttaeeag 4080 tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 4140 55 cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc 4200 gaacgaccta caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc 4260 ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca 4320 cgagggaget tecaggggga aacgeetggt atetttatag teetgteggg tttegeeace 4380 tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg 4440 60 ccagcaacgc ggccttttta cggttcctgg ccttttgctg gccttttgct cacatgttct 4500 ttcctgcgtt atcccctgat tctgtggata accgtattac cgcctttgag tgagctgata 4560 ccgctcgccg cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaagagc 4620 gcccaatacg caaaccgcct ctccccgcgc gttggccgat tcattaatgc agctggcacg 4680 acaggtttcc cgactggaaa gcgggcagtg agcgcaacgc aattaatgtg agttagctca 4740 65 ctcattaggc accccaggct ttacacttta tgcttccggc tcgtatgttg tgtggaattg 4800 tgagcggata acaatttcac acaggaaaca gctatgacca tgattacgcc aagctagccc 4860 gggctagctt gcatgcctgc aggttt 4886

```
<210> 75
      <211> 4905
      <212> DNA
      <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: vector
10
           pCMV-SSV(NNLS)
      <400> 75
      aaacagtccg atgtacgggc cagatatacg cgttgacatt gattattgac tagttattaa 60
      tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg cgttacataa 120
15
     cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt gacgtcaata 180
      atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca atgggtggac 240
     tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc aagtacgccc 300
     cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta catgacctta 360
     tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac catggtgatg 420
20
     cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg atttccaagt 480
     ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg ggactttcca 540
     aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt acggtgggag 600
     gtctatataa gcagagctct ctggctaact agagaaccca ctgcttactg gcttatcgaa 660
     attaatacga ctcactatag ggagacccaa gctgactcta gacttaatta agcgttgggg 720
25
     tgagtactcc ctctcaaaag cgggcatgac ttctgcgcta agattgtcag tttccaaaaa 780
     cgaggaggat ttgatattca cctggcccgc ggtgatgcct ttgagggtgg ccgcgtccat 840
     ctggtcagaa aagacaatct ttttgttgtc aagcttgagg tgtggcaggc ttgagatctg 900
     gccatacact tgagtgacat tgacatccac tttgcctttc tctccacagg tgtccactcc 960
     cagggcggcc gcaccatgcc caagaagaag aggaaggtga cgaaagataa gacgcgttat 1020
30
     aaatacgggg attatatttt acgcgagagg aaagggcggt attatgttta caagctagag 1080
     tatgaaaacg gtgaggtaaa agagcgttac gtgggtcctt tagctgacgt cgttgaatca 1140
     tatctaaaaa tgaaattagg ggtcgtaggg gatactcccc tacaagcgga tccccccggt 1200
     ttcgagcccg ggacaagcgg aagcggtggt ggaaaagagg gaactgaacg acgtaaaata 1260
     gegttggttg ccaatttgeg ccaataegeg aeggaeggea acataaagge gttetacaae 1320
35
     tatctcatga acgaaagggg gataagcgaa aaaactgcaa aggactacat caatgctata 1380
     tcaaagccgt ataaagagac gagagacgca cagaaggctt accgactctt tgcacgtttc 1440
     ttagcgtcac gcaatatcat acatgatgaa tttgcggata aaatattgaa agcggtaaag 1500
     gtgaagaagg cgaacgctga tatctacatt ccaacgttgg aagagataaa aaggacgtta 1560
     caattagcaa aagactatag cgaaaacgtc tacttcatct accgtatcgc tctcgagtcg 1620 ggcgttaggc tgagcgaaat actgaaagtg ctgaaggaac ccgaaaggga catttgcggt 1680
40
     aacgacgtct gttattatcc gcttagttgg actaggggat ataagggcgt cttctatgta 1740
     ttccacataa cgcctctgaa gagagtagag gtgacgaagt gggcaatagc ggactttgaa 1800
     cgacgtcata aggacgctat agcgataaag tacttccgca aattcgtagc gtctaagatg 1860
     gctgagctaa gcgtaccgtt agatattatc gattttattc aagggcgtaa accgacacgc 1920
45
     gttttaacgc aacattacgt atcgctcttc ggcatagcga aagagcaata taaaaagtat 1980
     gcggaatggc taaaaggggt ctgactcgag ggggggcccg tcgacctcga gatccaggcg 2040
     cggatcaata aaagatcatt attttcaata gatctgtgtg ttggtttttt gtgtgccttg 2100
     ggggagggg aggccagaat gaggcgcggc caagggggag ggggaggcca gaatgacctt 2160
     gggggagggg gaggccagaa tgaccttggg ggagggggag gccagaatga ggcgcgcccc 2220
50
     cgggtaccga gctcgaattc actggccgtc gttttacaac gtcgtgactg ggaaaaccct 2280 ggcgttaccc aacttaatcg ccttgcagca catccccctt tcgccagctg gcgtaatagc 2340
     gaagaggccc gcaccgatcg cccttcccaa cagttgcgca gcctgaatgg cgaatggcgc 2400
     ctgatgcggt attttctcct tacgcatctg tgcggtattt cacaccgcat atggtgcact 2460
     ctcagtacaa tetgetetga tgccgcatag ttaagccage eecgacacee gecaacacee 2520
55
     gctgacgcgc cctgacgggc ttgtctgctc ccggcatccg cttacagaca agctgtgacc 2580
     gtctccggga gctgcatgtg tcagaggttt tcaccgtcat caccgaaacg cgcgagacga 2640
     aagggcctcg tgatacgcct attttatag gttaatgtca tgataataat ggtttcttag 2700
     acgtcaggtg gcacttttcg gggaaatgtg cgcggaaccc ctatttgttt atttttctaa 2760
     atacattcaa atatgtatcc gctcatgaga caataaccct gataaatgct tcaataatat 2820
60
     tgaaaaagga agagtatgag tattcaacat ttccgtgtcg cccttattcc cttttttgcg 2880
     gcattttgcc ttcctgtttt tgctcaccca gaaacgctgg tgaaagtaaa agatgctgaa 2940
     gatcagttgg gtgcacgagt gggttacatc gaactggatc tcaacagcgg taagatcctt 3000
     gagagttttc gccccgaaga acgttttcca atgatgagca cttttaaagt tctgctatgt 3060
     ggcgcggtat tatcccgtat tgacgccggg caagagcaac tcggtcgccg catacactat 3120
65
     teteagaatg aettggttga gtacteacca gteacagaaa ageatettae ggatggeatg 3180
     acagtaagag aattatgcag tgctgccata accatgagtg ataacactgc ggccaactta 3240
     cttctgacaa cgatcggagg accgaaggag ctaaccgctt ttttgcacaa catgggggat 3300
```

```
catgtaactc gccttgatcg ttgggaaccg gagctgaatg aagccatacc aaacgacgag 3360
      cgtgacacca cgatgcctgt agcaatggca acaacgttgc gcaaactatt aactggcgaa 3420
      ctacttactc tagetteecg gcaacaatta atagactgga tggaggegga taaagttgca 3480
      ggaccactte tgegetegge cetteegget ggetggttta ttgetgataa atetggagee 3540
      ggtgagcgtg ggtctcgcgg tatcattgca gcactggggc cagatggtaa gccctcccgt 3600
      atcytagtta tctacacgac ggggagtcag gcaactatgg atgaacgaaa tagacagatc 3660
      gctgagatag gtgcctcact gattaagcat tggtaactgt cagaccaagt ttactcatat 3720
      atactitaga tigatttaaa acticatitt taatttaaaa ggatctaggt gaagatcett 3780
      tttgataatc tcatgaccaa aatcccttaa cgtgagtttt cgttccactg agcgtcagac 3840
      cccgtagaaa agatcaaagg atcttcttga gatccttttt ttctgcgcgt aatctgctgc 3900
      ttgcaaacaa aaaaaccacc getaccagcg gtggtttgtt tgccggatca agagctacca 3960
      actettttte egaaggtaae tggetteage agagegeaga taccaaatae tgteetteta 4020
      gtgtagccgt agttaggcca ccacttcaag aactctgtag caccgcctac atacctcgct 4080
      ctgctaatcc tgttaccagt ggctgctgcc agtggcgata agtcgtgtct taccgggttg 4140
 15
      gactcaagac gatagttacc ggataaggcg cagcggtcgg gctgaacggg gggttcgtgc 4200
      acacagecea gettggageg aacgaectae acegaactga gatacetaea gegtgageta 4260
      tgagaaagcg ccacgcttcc cgaagggaga aaggcggaca ggtatccggt aagcggcagg 4320
      gtcggaacag gagagcgcac gagggagctt ccagggggaa acgcctggta tctttatagt 4380
      cctgtcgggt ttcgccacct ctgacttgag cgtcgatttt tgtgatgctc gtcagggggg 4440
20
     cggagcctat ggaaaaacgc cagcaacgcg gcctttttac ggttcctggc cttttgctgg 4500
     cettttgctc acatgttctt teetgegtta teecetgatt etgtggataa eegtattace 4560
     geetttgagt gagetgatae egetegeege ageegaacga eegagegeag egagteagtg 4620
     agegaggaag eggaagageg eccaataege aaacegeete teeegegeg ttggeegatt 4680
     cattaatgca getggcacga caggttteec gaetggaaag egggcagtga gegcaaegea 4740 attaatgtga gttageteae teattaggea ecceaggett tacaetttat getteegget 4800
      cgtatgttgt gtggaattgt gagcggataa caatttcaca caggaaacag ctatgaccat 4860
      gattacgcca agctagcccg ggctagcttg catgcctgca ggttt
                                                                             4905
30
     <210> 76
     <211> 5290
      <212> DNA
     <213> Artificial Sequence
35
     <220>
     <223> Description of Artificial Sequence: vector
            pCMVXisA
     <400> 76
40
     agtccgatgt acgggccaga tatacgcgtt gacattgatt attgactagt tattaatagt 60
     aatcaattac ggggtcatta gttcatagcc catatatgga gttccgcgtt acataactta 120
     cggtaaatgg cccgcctggc tgaccgccca acgacccccg cccattgacg tcaataatga 180
     cgtatgttcc catagtaacg ccaataggga ctttccattg acgtcaatgg gtggactatt 240 tacggtaaac tgcccacttg gcagtacatc aagtgtatca tatgccaagt acgcccccta 300
     ttgacgtcaa tgacggtaaa tggcccgcct ggcattatgc ccagtacatg accttatggg 360
     actttcctac ttggcagtac atctacgtat tagtcatcgc tattaccatg gtgatgcggt 420
     tttggcagta catcaatggg cgtggatagc ggtttgactc acggggattt ccaagtctcc 480 accccattga cgtcaatggg agtttgttt ggcaccaaaa tcaacgggac tttccaaaat 540
     gtcgtaacaa ctccgcccca ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct 600
50
     atataagcag agetetetgg ctaactagag aacceactge ttactggett ategaaatta 660
     atacgactca ctatagggag acccaagctg actctagact taattaagcg ttggggtgag 720
     tactccctct caaaagcggg catgacttct gcgctaagat tgtcagtttc caaaaacgag 780
     gaggatttga tattcacctg gcccgcggtg atgcctttga gggtggccgc gtccatctgg 840 tcagaaaaga caatctttt gttgtcaagc ttgaggtgtg gcaggcttga gatctggcca 900
55
     tacacttgag tgacattgac atccactttg cctttctctc cacaggtgtc cactcccagg 960
     geggeegeee gatatgeaaa ateagggtea agacaaatat caacaageet ttgeagaett 1020
     agagecaete teatetaceg aeggeagttt teteggetea agtetgeaag cacageagea 1080
     aagagaacac atgagaacaa aagtactaca agacctagac aaggtaaatc tgcgtttgaa 1140
     gtctgcaaag acgaaagtct cagttcgaga atctaacgga agtctgcaat tacgagcaac 1200
     gttaccaatt aaacctggag ataaggacac caacggtaca ggcagaaagc aatacaatct 1260
     cagettgaat atecetgeaa aettggatgg aetgaagaeg getgaggaag aagettatga 1320 attaggtaaa ttaategete ggaaaacett tgaatggaat gataaatatt taggeaaaga 1380
     agccactaaa aaagattcac aaacaatagg tgatttacta gaaaaatttg cagaagagta 1440
     ttttaaaacc cataaacgca ccactaaaag cgaacatacc tttttttact atttttcccg 1500
65
     tgagcaaatc gataaagaat gggcgagata taatgccgcc agagccatat cagctttttg 1620
     cataacattc aatatagaaa ttgatttgtc ccagtattcc aaaatgcctg atcgcaattc 1680
```

```
gcgcaacatc cccacagatg cagaaatact atcaggaatt accaaatttg aagactatct 1740
     agttaccaga ggaaatcaag ttaatgaaga tgtaaaagat agctggcaac tttggcgctg 1800
     gacatatgga atgttagcag tttttggttt acgccccagg gaaattttta ttaaccctaa 1860
     tattgattgg tggttaagca aagagaatat agacctcaca tggaaagtag acaaagaatg 1920
     taaaactggt gaaagacaag cattaccctt acataaagaa tggattgatg agtttgattt 1980
     aagaaatccg aaatatttag aaatgctggc aacagcaatt agtaaaaaag ataaaacaaa 2040
     tcatgctgaa ataacagcct taactcagcg tattagttgg tggtttcgga aagtcgaatt 2100
     agattttaaa ccctatgatt tacgtcacgc ctgggcaatt agagcgcata ttttaggcat 2160
     accaatcaaa gcggcggctg ataatttggg gcatagtatg caggttcata cacaaaccta 2220
10
     tcagcgctgg ttctcgctag atatgcggaa gttagcgatt aatcaggctt tgactaagag 2280
     gaatgaatti gaggtgatta gggaggagaa tgctaaattg cagatagaaa atgaaaggtt 2340
     gaggatggaa attgagaagt taaagatgga aatagcttat aagaatagtt gagcggccgc 2400
     gtcgacctcg agatccaggc gcggatcaat aaaagatcat tattttcaat agatctgtgt 2460
     gttggttttt tgtgtgcctt gggggagggg gaggccagaa tgaggcgcgg ccaaggggga 2520
15
     gggggaggcc agaatgacct tgggggaggg ggaggccaga atgaccttgg gggaggggga 2580 ggccagaatg aggcgcgccc ccgggtaccg agctcgaatt cactggccgt cgttttacaa 2640
     cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgcagc acatccccct 2700
     ttcgccagct ggcgtaatag cgaagaggcc cgcaccgate gcccttccca acagttgcgc 2760
     agcctgaatg gegaatggeg cetgatgegg tattttetec ttacgcatet gtgcggtatt 2820
20
     tcacaccgca tatggtgcac tctcagtaca atctgctctg atgccgcata gttaagccag 2880
     ccccgacacc cgccaacacc cgctgacgcg ccctgacggg cttgtctgct cccggcatcc 2940
     gettacagae aagetgtgae egteteeggg agetgeatgt gteagaggtt tteaeegtea 3000
     tcaccgaaac gcgcgagacg aaagggcctc gtgatacgcc tatttttata ggttaatgtc 3060
     atgataataa tggtttctta gacgtcaggt ggcacttttc ggggaaatgt gcgcggaacc 3120
25
     cctatttgtt tatttttcta aatacattca aatatgtatc cgctcatgag acaataaccc 3180
     tgataaatgc ttcaataata ttgaaaaagg aagagtatga gtattcaaca tttccgtgtc 3240
     gcccttattc ccttttttgc ggcattttgc cttcctgttt ttgctcaccc agaaacgctg 3300
     gtgaaagtaa aagatgctga agatcagttg ggtgcacgag tgggttacat cgaactggat 3360 ctcaacagcg gtaagatcct tgagagttt cgccccgaag aacgttttcc aatgatgagc 3420
30
     acttttaaag ttctgctatg tggcgcggta ttatcccgta ttgacgccgg gcaagagcaa 3480
     cteggtegee geatacacta tteteagaat gaettggttg agtacteace agteacagaa 3540
     aagcatctta cggatggcat gacagtaaga gaattatgca gtgctgccat aaccatgagt 3600
     gataacactg cggccaactt acttctgaca acgatcggag gaccgaagga gctaaccgct 3660
     tttttgcaca acatggggga tcatgtaact cgccttgatc gttggggaacc ggagctgaat 3720
35
     gaagccatac caaacgacga gcgtgacacc acgatgcctg tagcaatggc aacaacgttg 3780
     cqcaaactat taactqqcqa actacttact ctaqcttccc qqcaacaatt aataqactqq 3840
     atggaggcgg ataaagttgc aggaccactt ctgcgctcgg cccttccggc tggctggttt 3900
     attgctgata aatctggagc cggtgagcgt gggtctcgcg gtatcattgc agcactgggg 3960
     ccagatggta agccctcccg tatcgtagtt atctacacga cggggagtca ggcaactatg 4020 gatgaacgaa atagacagat cgctgagata ggtgcctcac tgattaagca ttggtaactg 4080
40
     tcagaccaag tttactcata tatactttag attgatttaa aacttcattt ttaatttaaa 4140
     aggatctagg tgaagatcct ttttgataat ctcatgacca aaatccctta acgtgagttt 4200
     togttocact gagogtoaga coccgtagaa aagatoaaag gatottottg agatoctttt 4260
     tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac cgctaccagc ggtggtttgt 4320
45
     ttgccggatc aagagctacc aactcttttt ccgaaggtaa ctggcttcag cagagcgcag 4380
     ataccaaata ctgtccttct agtgtagccg tagttaggcc accacttcaa gaactctgta 4440
     gcaccgccta catacctcgc tctgctaatc ctgttaccag tggctgctgc cagtggcgat 4500
     aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac cggataaggc gcagcggtcg 4560
     ggctgaacgg ggggttcgtg cacacagccc agcttggagc gaacgaccta caccgaactg 4620
50
     agatacetac agegtgaget atgagaaage gecaegette eegaagggag aaaggeggac 4680
     aggtateegg taageggeag ggteggaaca ggagagegea egagggaget teeaggggga 4740 aacgeetggt atetttatag teetgteggg tttegeeace tetgaettga gegtegattt 4800
     ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg ccagcaacgc ggccttttta 4860
     cggttcctgg ccttttgctg gccttttgct cacatgttct ttcctgcgtt atcccctgat 4920
55
     tetgtggata accgtattac cgcetttgag tgagetgata cegetegeeg cageegaacg 4980
     accgagegca gegagteagt gagegaggaa geggaagage geccaataeg caaacegeet 5040
     etcecegege gttggeegat teattaatge agetggeaeg acaggtttee egactggaaa 5100
     gcgggcagtg agcgcaacgc aattaatgtg agttagctca ctcattaggc accccaggct 5160
     ttacactita tgcttccggc tcgtatgitg tgtggaattg tgagcggata acaatttcac 5220
60
     acaggaaaca gctatgacca tgattacgcc aagctagccc gggctagctt gcatgcctgc 5280
     aggtttaaac
```

<213> Artificial Sequence

<210> 77 65 <211> 5309 <212> DNA

<220>

```
<223> Description of Artificial Sequence: vector
           pCMVXisANNLS
     <400> 77
     aaacagtccg atgtacgggc cagatatacg cgttgacatt gattattgac tagttattaa 60
     tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg cgttacataa 120
     cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt gacgtcaata 180
10
     atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca atgggtggac 240
     tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc aagtacgccc 300
     cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta catgacctta 360
     tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac catggtgatg 420
     cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg atttccaagt 480
15
     ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg ggactttcca 540
     aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt acggtgggag 600
     gtctatataa gcagagetet etggetaaet agagaaeeea etgettaetg gettategaa 660
     attaatacga ctcactatag ggagacccaa gctgactcta gacttaatta agcgttgggg 720
     tgagtactee eteteaaaag egggeatgae ttetgegeta agattgteag ttteeaaaaa 780
20
     cgaggaggat ttgatattca cctggcccgc ggtgatgcct ttgagggtgg ccgcgtccat 840
     ctggtcagaa aagacaatct ttttgttgtc aagcttgagg tgtggcaggc ttgagatctg 900
     gocatacaet tgagtgacat tgacatecae tttgeettte tetecaeagg tgtecaetee 960
     cagggcggcc gcaccatgcc caagaagaag aggaaggtgc aaaatcaggg tcaagacaaa 1020
     tatcaacaag cctttgcaga cttagagcca ctttcatcta ccgacggcag ttttctcggc 1080
25
     tcaagtetge aagcacagca gcaaagagaa cacatgagaa caaaagtact acaagaccta 1140 -
     gacaaggtaa atctgcgttt gaagtctgca aagacgaaag tctcagttcg agaatctaac 1200
     ggaagtctgc aattacgagc aacgttacca attaaacctg gagataagga caccaacggt 1260
     acaggeagaa agcaatacaa teteagettg aatateeetg caaacttgga tggactgaag 1320
     acggctgagg aagaagctta tgaattaggt aaattaatcg ctcggaaaac ctttgaatgg 1380
30
     aatgataaat atttaggcaa agaagccact aaaaaaagatt cacaaacaat aggtgattta 1440
     ctagaaaaat ttgcagaaga gtattttaaa acccataaac gcaccactaa aagcgaacat 1500
     acctttttt actattttc ccgcacccaa cgatatacca attccaaaga tttagcaacg 1560
     gcggaaaatc tcatcaattc aattgagcaa atcgataaag aatgggcgag atataatgcc 1620
     gccagagcca tatcagcttt ttgcataaca ttcaatatag aaattgattt gtcccagtat 1680
35
     tccaaaatgc ctgatcgcaa ttcgcgcaac atccccacag atgcagaaat actatcagga 1740
     attaccaaat ttgaagacta tctagttacc agaggaaatc aagttaatga agatgtaaaa 1800
     gatagctggc aactttggcg ctggacatat ggaatgttag cagtttttgg tttacgcccc 1860
     agggaaattt ttattaaccc taatattgat tggtggttaa gcaaagagaa tatagacctc 1920
     acatggaaag tagacaaaga atgtaaaact ggtgaaagac aagcattacc cttacataaa 1980
40
     gaatggattg atgagtttga tttaagaaat ccgaaatatt tagaaatgct ggcaacagca 2040
     attagtaaaa aagataaaac aaatcatgct gaaataacag ccttaactca gcgtattagt 2100
     tggtggtttc ggaaagtcga attagatttt aaaccctatg atttacgtca cgcctgggca 2160
     atcagagege atattttagg cataccaatc aaageggegg ctgataattt ggggcatagt 2220 atgeaggtte atacacaaac ctatcagege tggttetege tagatatgeg gaagttageg 2280
     attaatcagg ctttgactaa gaggaatgaa tttgaggtga ttagggagga gaatgctaaa 2340
     ttgcagatag aaaatgaaag gttgaggatg gaaattgaga agttaaagat ggaaatagct 2400
     tataagaata gttgagcggc cgcgtcgacc tcgagatcca ggcgcggatc aataaaagat 2460
     cattattttc aatagatctg tgtgttggtt ttttgtgtgc cttgggggag ggggaggcca 2520
     gaatgaggcg cggccaaggg ggagggggag gccagaatga ccttggggga gggggaggcc 2580
50
     agaatgacct tgggggaggg ggaggccaga atgaggcgcg cccccgggta ccgagctcga 2640
     attcactggc cgtcgtttta caacgtcgtg actgggaaaa ccctggcgtt acccaactta 2700
     ategeettge ageacatece cetttegeea getggegtaa tagegaagag geeegeaceg 2760
     atcgcccttc ccaacagttg cgcagcctga atggcgaatg gcgcctgatg cggtattttc 2820
     teettaegea tetgtgeggt attteaeace geatatggtg eacteteagt acaatetget 2880
55
     ctgatgccgc atagttaagc cagccccgac acccgccaac acccgctgac gcgccctgac 2940
     gggcttgtct gctcccggca tccgcttaca gacaagctgt gaccgtctcc gggagctgca 3000
     tgtgtcagag gttttcaccg tcatcaccga aacgcgcgag acgaaagggc ctcgtgatac 3060
     gcctatttt ataggttaat gtcatgataa taatggtttc ttagacgtca ggtggcactt 3120
     ttcggggaaa tgtgcgcgga acccctattt gtttatttt ctaaatacat tcaaatatgt 3180
60
     atcegeteat gagacaataa eeetgataaa tgetteaata atattgaaaa aggaagagta 3240
     tgagtattca acatttccgt gtcgccctta ttcccttttt tgcggcattt tgccttcctg 3300
     tttttgctca cccagaaacg ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac 3360
     gagtgggtta catcgaactg gatctcaaca gcggtaagat ccttgagagt tttcgccccg 3420
     aagaacgttt tccaatgatg agcactttta aagttctgct atgtggcgcg gtattatccc 3480
65
     gtattgacgc cgggcaagag caactcggtc gccgcataca ctattctcag aatgacttgg 3540
     ttgagtactc accagtcaca gaaaagcatc ttacggatgg catgacagta agagaattat 3600
     gcagtgctgc cataaccatg agtgataaca ctgcggccaa cttacttctg acaacgatcg 3660
```

```
68
     gaggaccgaa ggagctaacc gcttttttgc acaacatggg ggatcatgta actcgccttg 3720
     atcgttggga accggagctg aatgaagcca taccaaacga cgagcgtgac accacgatgc 3780
     ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagctt 3840
     cccggcaaca attaatagac tggatggagg cggataaagt tgcaggacca cttctgcgct 3900
     cggcccttcc ggctggctgg tttattgctg ataaatctgg agccggtgag cgtgggtctc 3960
     gcggtatcat tgcagcactg gggccagatg gtaagccctc ccgtatcgta gttatctaca 4020 cgacggggag tcaggcaact atggatgaac gaaatagaca gatcgctgag ataggtgcct 4080
     cactgattaa gcattggtaa ctgtcagacc aagtttactc atatatactt tagattgatt 4140
     taaaacttca tttttaattt aaaaggatct aggtgaagat cctttttgat aatctcatga 4200
10
     ccaaaatccc ttaacgtgag ttttcgttcc actgagcgtc agaccccgta gaaaagatca 4260
     aaggatette ttgagateet tttttetge gegtaatetg etgettgeaa acaaaaaaaa 4320
     caccgctacc agcggtggtt tgtttgccgg atcaagagct accaactctt tttccgaagg 4380
     taactggett cagcagagcg cagataccaa atactgteet tetagtgtag ccgtagttag 4440
     gccaccactt caagaactet gtagcaccgc ctacatacet cgctctgcta atcctgttac 4500
15
     cagtggctgc tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt 4560
     taccggataa ggcgcagcgg tcgggctgaa cggggggttc gtgcacacag cccagcttgg 4620
     agcgaacgac ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcgccacgc 4680
     ttcccgaagg gagaaaggcg gacaggtatc cggtaagcgg cagggtcgga acaggagagc 4740
     gcacgaggga gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc gggtttcgcc 4800
20
     acctetgact tgagegtega tttttgtgat getegteagg ggggeggage ctatggaaaa 4860
     acgccagcaa cgcggccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt 4920
     tettteetge gttateeeet gattetgtgg ataacegtat taeegeettt gagtgagetg 4980
     ataccgeteg cegeageega aegaeegage geagegagte agtgagegag gaageggaag 5040
     agegeceaat aegeaaaceg ceteteceeg egegttggee gatteattaa tgeagetgge 5100
     acgacaggtt tcccgactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc 5160
     tcactcatta ggcaccccag getttacact ttatgettee ggetegtatg ttgtgtggaa 5220
     ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac gccaagctag 5280
     cccgggctag cttgcatgcc tgcaggttt
30
     <210> 78
     <211> 7608
     <212> DNA
     <213> Artificial Sequence
35
     <220>
     <223> Description of Artificial Sequence: vector
           pPGKnifD
40
     <400> 78
     tcgaggaatt ctaccgggta ggggaggcgc ttttcccaag gcagtctgga gcatgcgctt 60
     tagcagecce getgggeact tggegetaca caagtggeet etggetegea cacatteeac 120
     atccaccggt aggcgccaac cggctccgtt ctttggtggc cccttcgcgc caccttctac 180
     tecteceeta gteaggaagt tececeege ecegeagete gegtegtgea ggacgtgaca 240
45
     aatggaagta gcacgtctca ctagtctcgt gcagatggac agcaccgctg agcaatggaa 300
     gcgggtaggc ctttggggca gcggccaata gcagctttgc tccttcgctt tctgggctca 360
    getgttetee tetteeteat eteegggeet ttegaeetge ageeeggtae agttegaatg 540
50
    getetteeet teegteaaat geactettgg gattacteeg aacetagega tggggtgeaa 600
     atgtcagatc agataagttc gaataacttc gtatagcata cattatacga agttataagc 660
     ttgcatgcct gcaggtcggc cgccacgacc ggccggccgg tgccgccacc atcccctgac 720
     ccacgcccct gacccctcac aaggagacga ccttccatga ccgagtacaa gcccacggtg 780
     cgcctcgcca cccgcgacga cgtcccccgg gccgtacgca ccctcgccgc cgcgttcgcc 840
    gactacceeg ccacgegeca caccgtegac ceggacegec acategageg ggtcacegag 900
     ctgcaagaac tetteeteac gegegteggg etegacateg gcaaggtgtg ggtegeggae 960
     gacggcgceg cggtggcggt ctggaccacg ccggagagcg tcgaagcggg ggcggtgttc 1020
    gccgagatcg gcccgcgcat ggccgagttg agcggttccc ggctggccgc gcagcaacag 1080 atggaaggcc tcctggcgcc gcaccggccc aaggagcccg cgtggttcct ggccaccgtc 1140
60
    ggcgtctcgc ccgaccacca gggcaagggt ctgggcagcg ccgtcgtgct ccccggagtg 1200
    gaggeggeeg agegegegg ggtgeeegee tteetggaga ceteegegee eegeaacete 1260
    cccttctacg agcggctcgg cttcaccgtc accgccgacg tcgagtgccc gaaggaccgc 1320
     gcgacctggt gcatgacccg caagcccggt gcctgacgcc cgccccacga cccgcagcgc 1380
    ccgaccgaaa ggagcgcacg accccatggc tccgaccgaa gccgacccgg gcggcccgc 1440
    cgaccccgca cccgcccccg aggcccaccg actctagagg atcataatca gccataccac 1500
    atttgtagag gttttacttg ctttaaaaaa cctcccacac ctccccctga acctgaaaca 1560
     taaaatgaat gcaattgttg ttgttaactt gtttattgca gcttataatg gttacaaata 1620
```

aagcaatagc atcacaaatt tcacaaataa agcattttt tcactgcatt ctagttgtgg 1680 tttgtccaaa ctcatcaatg tatcttatca tgtctggatc cagctgttga aagctattaa 1740 accacaaaaa ggattactcc ggcccttatc acggttacga cggatttgga tccataactt 1800 cgtatagcat acattatacg aagttatacc gggccaccat ggtcgcgagt agcttggcac 1860 tggccgtcgt tttacaacgt cgtgactggg aaaaccctgg cgttacccaa cttaatcgcc 1920 ttgcagcaca tccccctttc gccagctggc gtaatagcga agaggcccgc accgatcgcc 1980 cttcccaaca gttgcgcagc ctgaatggcg aatggcgctt tgcctggttt ccggcaccag 2040 aagcggtgcc ggaaagctgg ctggagtgcg atcttcctga ggccgatact gtcgtcgtcc 2100 cctcaaactg gcagatgcac ggttacgatg cgcccatcta caccaacgta acctatccca 2160 10 ttacggtcaa tccgccgttt gttcccacgg agaatccgac gggttgttac tcgctcacat 2220 ttaatgttga tgaaagctgg ctacaggaag gccagacgcg aattatttt gatggcgtta 2280 acteggegtt teatetgtgg tgeaacggge getgggtegg ttacggecag gacagtegtt 2340 tgeegtetga atttgacetg agegeatttt tacgegeegg agaaaacege etegeggtga 2400 tggtgctgcg ttggagtgac ggcagttatc tggaagatca ggatatgtgg cggatgagcg 2460 15 gcattttccg tgacgtctcg ttgctgcata aaccgactac acaaatcagc gatttccatg 2520 ttgccactcg ctttaatgat gatttcagcc gcgctgtact ggaggctgaa gttcagatgt 2580 gcggcgagtt gcgtgactac ctacgggtaa cagtttcttt atggcagggt gaaacgcagg 2640 tegecagegg cacegegeet tteggeggtg aaattatega tgagegtggt ggttatgeeg 2700 atcgcgtcac actacgtctg aacgtcgaaa acccgaaact gtggagcgcc gaaatcccga 2760 20 atctctatcg tgcggtggtt gaactgcaca ccgccgacgg cacgctgatt gaagcagaag 2820 cctgcgatgt cggtttccgc gaggtgcgga ttgaaaatgg tctgctgctg ctgaacggca 2880 tggatgagca gacgatggtg caggatatcc tgctgatgaa gcagaacaac tttaacgccg 3000 tgcgctgttc gcattatccg aaccatccgc tgtggtacac gctgtgcgac cgctacggcc 3060 25 tgtatgtggt ggatgaagcc aatattgaaa cccacggcat ggtgccaatg aatcgtctga 3120 ccgatgatec gegetggeta ceggegatga gegaacgegt aacgegaatg gtgcagegeg 3180 ategtaatea eeegagtgtg ateatetggt egetggggaa tgaateagge caeggegeta 3240 atcacgacgc gctgtatcgc tggatcaaat ctgtcgatcc ttcccgcccg gtgcagtatg 3300 aaggeggegg ageegacace aeggeeaceg atattatttg eeegatgtae gegegegtgg 3360 30 atgaagacca gcccttcccg gctgtgccga aatggtccat caaaaaatgg ctttcgctac 3420 ctggagagac gcgcccgctg atcctttgcg aatacgccca cgcgatgggt aacagtcttg 3480 gcggtttcgc taaatactgg caggcgtttc gtcagtatcc ccgtttacag ggcggcttcg 3540 tetgggaetg ggtggateag tegetgatta aatatgatga aaaeggeaae eegtggtegg 3600 cttacggcgg tgattttggc gatacgccga acgatcgcca gttctgtatg aacggtctgg 3660 35 tetttgeega eegeaegeeg cateeagege tgaeggaage aaaacaceag eageagtttt 3720 tccagttccg tttatccggg caaaccatcg aagtgaccag cgaatacctg ttccgtcata 3780 gcgataacga gctcctgcac tggatggtgg cgctggatgg taagccgctg gcaagcggtg 3840 aagtgeetet ggatgteget eeacaaggta aacagttgat tgaactgeet gaactacege 3900 agccggagag cgccgggcaa ctctggctca cagtacgcgt agtgcaaccg aacgcgaccg 3960 40 catggtcaga agccgggcac atcagcgcct ggcagcagtg gcgtctggcg gaaaacctca 4020 gtgtgacget eccegeegeg teccaegeea tecegeatet gaecaecage gaaatggatt 4080 tttgeatega getgggtaat aagegttgge aatttaaceg ceagteagge tttettteae 4140 agatgtggat tggcgataaa aaacaactgc tgacgccgct gcgcgatcag ttcacccgtg 4200 caccgctgga taacgacatt ggcgtaagtg aagcgacccg cattgaccct aacgcctggg 4260 45 togaacgetg gaaggeggeg ggccattacc aggccgaagc agcgttgttg cagtgcacgg 4320 cagatacact tgctgatgcg gtgctgatta cgaccgctca cgcgtggcag catcagggga 4380 aaaccttatt tatcagccgg aaaacctacc ggattgatgg tagtggtcaa atggcgatta 4440 ccgttgatgt tgaagtggcg agcgatacac cgcatccggc gcggattggc ctgaactgcc 4500 agetggegea ggtageagag egggtaaaet ggeteggatt agggeegeaa gaaaaetate 4560 50 ecgaccgcct tactgccgcc tgttttgacc gctgggatct gccattgtca gacatgtata 4620 ccccgtacgt cttcccgagc gaaaacggtc tgcgctgcgg gacgcgcgaa ttgaattatg 4680 gcccacacca gtggcgcggc gacttccagt tcaacatcag ccgctacagt caacagcaac 4740 tgatggaaac cagccatcgc catctgctgc acgcggaaga aggcacatgg ctgaatatcg 4800 acggtttcca tatggggatt ggtggcgacg actcctggag cccgtcagta tcggcggaat 4860 55 tccagctgag cgccggtcgc taccattacc agttggtctg gtgtcaaaaa taataataac 4920 cgggcagggg ggatctttgt gaaggaacct tacttctgtg gtgtgacata attggacaaa 4980 ctacctacag agatttaaag ctctaaggta aatataaaat ttttaagtgt ataatgtgtt 5040 aaactactga ttctaattgt ttgtgtattt tagattccaa cctatggaac tgatgaatgg 5100 gagcagtggt ggaatgccag atccagacat gataagatac attgatgagt ttggacaaac 5160 cacaactaga atgcagtgaa aaaaatgctt tatttgtgaa atttgtgatg ctattgcttt 5220 60 atttgtaacc attataagct gcaataaaca agttaacaac aacaattgca ttcattttat 5280 gtttcaggtt cagggggagg tgtgggaggt tttttaaagc aagtaaaacc tctacaaatg 5340 tggtatggct gattatgatc tgcggccgca gggcctcgtg atacgcctat ttttataggt 5400 taatgtcatg ataataatgg tttcttagac gtcaggtggc acttttcggg gaaatgtgcg 5460 65 cggaacccct atttgtttat ttttctaaat acattcaaat atgtatccgc tcatgagaca 5520 ataaccctga taaatgcttc aataatattg aaaaaggaag agtatgagta ttcaacattt 5580 cogtgtogcc cttattccct tttttgcggc attttgcctt cctgtttttg ctcacccaga 5640

```
aacgctggtg aaagtaaaag atgctgaaga tcagttgggt gcacgagtgg gttacatcga 5700
     actggatete aacageggta agateettga gagttttege eeegaagaac gttttecaat 5760
     gatgagcact tttaaagttc tgctatgtgg cgcggtatta tcccgtattg acgccgggca 5820
     agagcaactc ggtcgccgca tacactattc tcagaatgac ttggttgagt actcaccagt 5880
     cacagaaaag catcttacgg atggcatgac agtaagagaa ttatgcagtg ctgccataac 5940
     catgagtgat aacactgcgg ccaacttact tctgacaacg atcggaggac cgaaggagct 6000
     aaccgctttt ttgcacaaca tgggggatca tgtaactcgc cttgatcgtt gggaaccgga 6060
     gctgaatgaa gccataccaa acgacgagcg tgacaccacg atgcctgtag caatggcaac 6120
     aacgttgcgc aaactattaa ctggcgaact acttactcta gcttcccggc aacaattaat 6180
10
     agactggatg gaggcggata aagttgcagg accacttctg cgctcggccc ttccggctgg 6240 ctggtttatt gctgataaat ctggagccgg tgagcgtggg tctcgcggta tcattgcagc 6300
     actggggcca gatggtaagc cotcocgtat cgtagttatc tacacgacgg ggagtcaggc 6360
     aactatggat gaacgaaata gacagatcgc tgagataggt gcctcactga ttaagcattg 6420
     gtaactgtca gaccaagttt actcatatat actttagatt gatttaaaac ttcattttta 6480
15
     atttaaaagg atctaggtga agatcctttt tgataatctc atgaccaaaa tcccttaacg 6540
     tgagttttcg ttccactgag cgtcagaccc cgtagaaaag atcaaaggat cttcttgaga 6600
     tcctttttt ctgcgcgtaa tctgctgctt gcaaacaaaa aaaccaccgc taccagcggt 6660
     ggtttgtttg ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag 6720
     agcgcagata ccaaatactg tccttctagt gtagccgtag ttaggccacc acttcaagaa 6780
20
     ctctgtagca ccgcctacat acctcgctct gctaatcctg ttaccagtgg ctgctgccag 6840
     tggcgataag tcgtgtctta ccgggttgga ctcaagacga tagttaccgg ataaggcgca 6900
     gcggtcgggc tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac 6960
     cgaactgaga tacctacage gtgagetatg agaaagegee aegetteeeg aagggagaaa 7020
     ggcggacagg tatccggtaa gcggcagggt cggaacagga gagcgcacga gggagcttcc 7080
25
     agggggaaac gcctggtatc tttatagtcc tgtcgggttt cgccacctct gacttgagcg 7140
     togatttttg tgatgotogt caggggggcg gagcotatgg aaaaacgcca gcaacgcggc 7200
     ctttttacgg ttcctggcct tttgctggcc ttttgctcac atgttctttc ctgcgttatc 7260
     ccctgattct gtggataacc gtattaccgc ctttgagtga gctgataccg ctcgccgcag 7320
     ccgaacgacc gagcgcagcg agtcagtgag cgaggaagcg gaagagcgcc caatacgcaa 7380
30
     accgcctctc cccgcgcgtt ggccgattca ttaatgcagc tggcacgaca ggtttcccga 7440
     ctggaaagcg ggcagtgagc gcaacgcaat taatgtgagt tagctcactc attaggcacc 7500
     ccaggettta cactttatgc ttccggctcg tatgttgtgt ggaattgtga gcggataaca 7560
     atttcacaca ggaaacagct atgaccatga ttacgccaag ctggcgcg
35
     <210> 79
     <211> 7523
     <212> DNA
     <213> Artificial Sequence
40
     <220>
     <223> Description of Artificial Sequence: vector
           pPGKnifD3'
45
     <400> 79
     tcgaggaatt ctaccgggta ggggaggcgc ttttcccaag gcagtctgga gcatgcgctt 60
     tagcagecce getgggeact tggegetaca caagtggeet etggetegea cacattecae 120
     atecaceggt aggegeeaac eggeteegtt etttggtgge eeettegege eacettetae 180
     tectececta greaggaagt tecececege eeegeagete gegtegtgea ggaegtgaca 240
50
     aatggaagta gcacgtetca ctagtetegt gcagatggac agcaccgetg agcaatggaa 300
     gcgggtaggc ctttggggca gcggccaata gcagctttgc tccttcgctt tctgggctca 360
     gaggetggga aggggtgggt ccgggggcgg getcaggggc gggetcaggg gcggggcggg 420
     cgcccgaagg tcctccggag gcccggcatt ctgcacgctt caaaagcgca cgtctgccgc 480
     gctgttctcc tcttcctcat ctccgggcct ttcgacctgc agcccggtac agttcgaata 540
55
     acttcgtata gcatacatta tacgaagtta taagcttgca tgcctgcagg tcggccgcca 600
     cgaccggccg gccggtgccg ccaccatccc ctgacccacg cccctgaccc ctcacaagga 660
     gacgaccttc catgaccgag tacaagccca cggtgcgcct cgccacccgc gacgacgtcc 720
     cccgggccgt acgcaccctc gccgccgcgt tcgccgacta ccccgccacg cgccacaccg 780
     tegaceegga eegecacate gagegggtea eegagetgea agaactette etcacegegeg 840
60
     tcgggctcga catcggcaag gtgtgggtcg cggacgacgg cgccgcggtg gcggtctgga 900
     ccacgccgga gagcgtcgaa gcgggggcgg tgttcgccga gatcggcccg cgcatggccg 960
     agttgagegg ttcccggctg gccgcgcagc aacagatgga aggcctcctg gcgccgcacc 1020
     ggcccaagga gcccgcgtgg ttcctggcca ccgtcggcgt ctcgcccgac caccagggca 1080
     agggtctggg cagcgccgtc gtgctccccg gagtggaggc ggccgagcgc gccggggtgc 1140
65
     cogecttect ggagacetec gegeccegea aceteceett ctacgagegg cteggettea 1200
     ccgtcaccgc cgacgtcgag tgcccgaagg accgcgcgac ctggtgcatg acccgcaagc 1260
     ecggtgeetg acgeecgee cacgaecege agegeecgae egaaaggage geacgaecee 1320
```

WO 02/38613

```
atggeteega eegaageega eeegggegge eeegeegaee eegeaceege eeeegaggee 1380
     caccgactct agaggatcat aatcagccat accacatttg tagaggtttt acttgcttta 1440
     aaaaacctcc cacacctccc cctgaacctg aaacataaaa tgaatgcaat tqttqttqtt 1500
     aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 1560
     aataaagcat ttttttcact gcattctagt tgtggtttgt ccaaactcat caatgtatct 1620
     tatcatgtct ggatccagct gttgaaagct attaaaccac aaaaaggatt actccggccc 1680
     ttatcacggt tacgacggat ttggatccat aacttcgtat agcatacatt atacgaagtt 1740
     ataccgggcc accatggtcg cgagtagctt ggcactggcc gtcgttttac aacgtcgtga 1800
     ctgggaaaac cctggcgtta cccaacttaa tcgccttgca gcacatcccc ctttcgccag 1860
10
     ctggcgtaat agcgaagagg cccgcaccga tcgcccttcc caacagttgc gcagcctgaa 1920
     tggcgaatgg cgctttgcct ggtttccggc accagaagcg gtgccggaaa gctggctgga 1980
     gtgcgatctt cctgaggccg atactgtcgt cgtcccctca aactggcaga tgcacggtta 2040
     cgatgcgccc atctacacca acgtaaccta tcccattacg gtcaatccgc cgtttgttcc 2100
     cacggagaat ccgacgggtt gttactcgct cacatttaat gttgatgaaa gctggctaca 2160
     ggaaggccag acgcgaatta tttttgatgg cgttaactcg gcgtttcatc tgtggtgcaa 2220
     egggegetgg gteggttaeg geeaggaeag tegtttgeeg tetgaatttg acetgagege 2280
     atttttacgc gccggagaaa accgcctcgc ggtgatggtg ctgcgttgga gtgacggcag 2340
     ttatctggaa gatcaggata tgtggcggat gagcggcatt ttccgtgacg tctcgttgct 2400
     gcataaaccg actacacaaa tcagcgattt ccatgttgcc actcgcttta atgatgattt 2460.
.20
     cagcogogot gtactggagg ctgaagttca gatgtgcggc gagttgcgtg actacctacg 2520
     ggtaacagtt tetttatgge agggtgaaae geaggtegee ageggeaeeg egeetttegg 2580
     cggtgaaatt atcgatgagc gtggtggtta tgccgatcgc gtcacactac gtctgaacgt 2640
     cgaaaacccg aaactgtgga gcgccgaaat cccgaatctc tatcgtgcgg tggttgaact 2700
     geacacegee gacggeacge tgattgaage agaageetge gatgteggtt teegegaggt 2760
25
     geggattgaa aatggtetge tgetgetgaa eggeaageeg ttgetgatte gaggegttaa 2820
     cegteaegag cateatecte tgeatggtea ggteatggat gageagaega tggtgeagga 2880
     tatcctgctg atgaagcaga acaactttaa cgccgtgcgc tgttcgcatt atccgaacca 2940
     tccgctgtgg tacacgctgt gcgaccgcta cggcctgtat gtggtggatg aagccaatat 3000
     tgaaacccac ggcatggtgc caatgaatcg tctgaccgat gatccgcgct ggctaccggc 3060
30
     gatgagcgaa cgcgtaacgc gaatggtgca gcgcgatcgt aatcacccga gtgtgatcat 3120
     ctggtcgctg gggaatgaat caggccacgg cgctaatcac gacgcgctgt atcgctggat 3180
     caaatetgte gateetteec geeeggtgea gtatgaagge ggeggageeg acaecaegge 3240
     caccgatatt atttgcccga tgtacgcgcg cgtggatgaa gaccagccct tcccggctgt 3300
     gccgaaatgg tccatcaaaa aatggctttc gctacctgga gagacgcgcc cgctgatcct 3360
     ttgcgaatac gcccacgcga tgggtaacag tcttggcggt ttcgctaaat actggcaggc 3420
     gtttcgtcag tatccccgtt tacagggcgg cttcgtctgg gactgggtgg atcagtcgct 3480 gattaaatat gatgaaaacg gcaacccgtg gtcggcttac ggcggtgatt ttggcgatac 3540
     gccgaacgat cgccagttct gtatgaacgg tctggtcttt gccgaccgca cgccgcatcc 3600
     agcgctgacg gaagcaaaac accagcagca gtttttccag ttccgtttat ccgggcaaac 3660
     catcgaagtg accagcgaat acctgttccg tcatagcgat aacgagctcc tgcactggat 3720
     ggtggcgctg gatggtaagc cgctggcaag cggtgaagtg cctctggatg tcgctccaca 3780
     aggtaaacag ttgattgaac tgcctgaact accgcagccg gagagcgccg ggcaactctg 3840
     getcaeagta egegtagtge aacegaacge gacegeatgg teagaageeg ggeacateag 3900
     egectggeag cagtggegte tggeggaaaa ceteagtgtg aegeteeeeg eegegteeea 3960
45
     cgccatcccg catctgacca ccagcgaaat ggatttttgc atcgagctgg gtaataagcg 4020
     ttggcaattt aaccgccagt caggetttet ttcacagatg tggattggcg ataaaaaaca 4080
     actgctgacg ccgctgcgcg atcagttcac ccgtgcaccg ctggataacg acattggcgt 4140
     aagtgaagcg acccgcattg accctaacgc ctgggtcgaa cgctggaagg cggcgggcca 4200
     ttaccaggcc gaagcagcgt tgttgcagtg cacggcagat acacttgctg atgcggtgct 4260
50
     gattacgacc gctcacgcgt ggcagcatca ggggaaaacc ttatttatca gccggaaaac 4320
     ctaccggatt gatggtagtg gtcaaatggc gattaccgtt gatgttgaag tggcgagcga 4380
     tacaccgcat ccggcgcgga ttggcctgaa ctgccagctg gcgcaggtag cagagcgggt 4440
     aaactggctc ggattagggc cgcaagaaaa ctatcccgac cgccttactg ccgcctgttt 4500
     tgaccgctgg gatctgccat tgtcagacat gtataccccg tacgtcttcc cgagcgaaaa 4560
55
     eggtetgege tgegggaege gegaattgaa ttatggeeca caccagtgge geggegaett 4620
     ccagttcaac atcagccgct acagtcaaca gcaactgatg gaaaccagcc atcgccatct 4680
     gctgcacgcg gaagaaggca catggctgaa tatcgacggt ttccatatgg ggattggtgg 4740
     cgacgactcc tggagcccgt cagtatcggc ggaattccag ctgagcgccg gtcgctacca 4800
     ttaccagttg gtctggtgtc aaaaataata ataaccgggc aggggggatc tttgtgaagg 4860
60
     aacettactt ctgtggtgtg acataattgg acaaactacc tacagagatt taaageteta 4920
     aggtaaatat aaaattttta agtgtataat gtgttaaact actgattcta attgtttgtg 4980
     tattttagat tocaacctat ggaactgatg aatgggagca gtggtggaat gccagatcca 5040
     gacatgataa gatacattga tgagtttgga caaaccacaa ctagaatgca gtgaaaaaaa 5100
     tgctttattt gtgaaatttg tgatgctatt gctttatttg taaccattat aagctgcaat 5160
65
     aaacaagtta acaacaacaa ttgcattcat tttatgtttc aggttcaggg ggaggtgtgg 5220
     gaggtttttt aaagcaagta aaacctctac aaatgtggta tggctgatta tgatctgcgg 5280
     ccgcagggcc tcgtgatacg cctatttta taggttaatg toatgataat aatggtttct 5340
```

tagacgtcag gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatttttc 5400

```
taaatacatt caaatatgta teegeteatg agacaataae eetgataaat getteaataa 5460
     tattgaaaaa ggaagagtat gagtattcaa catttccgtg tcgcccttat tcccttttt 5520
     geggeatttt geetteetgt ttttgeteac eeagaaacge tggtgaaagt aaaagatget 5580
     gaagatcagt tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc 5640
     cttgagagtt ttcgccccga agaacgtttt ccaatgatga gcacttttaa agttctgcta 5700
     tgtggcgcgg tattatcccg tattgacgcc gggcaagagc aactcggtcg ccgcatacac 5760
     tatteteaga atgaettggt tgagtaetea ceagteacag aaaageatet taeggatgge 5820
     atgacagtaa gagaattatg cagtgctgcc ataaccatga gtgataacac tgcggccaac 5880
10
     ttacttctga caacgatcgg aggaccgaag gagctaaccg ctittttgca caacatgggg 5940
     gatcatgtaa ctcgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac 6000
     gagcgtgaca ccacgatgcc tgtagcaatg gcaacaacgt tgcgcaaact attaactggc 6060
     gaactactta etetagette eeggeaacaa ttaatagaet ggatggagge ggataaagtt 6120
     gcaggaccac ttctgcgctc ggcccttccg gctggctggt ttattgctga taaatctgga 6180 gccggtgagc gtgggtctcg cggtatcatt gcagcactgg ggccagatgg taagccctcc 6240
15
     cgtatcgtag ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag 6300
     atogotgaga taggtgooto actgattaag cattggtaac tgtcagacca agtttactca 6360
     tatatacttt agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc 6420
     ctttttgata atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca 6480
20
     gaccccgtag aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc 6540
     tgcttgcaaa caaaaaaacc accgctacca gcggtggttt gtttgccgga tcaagagcta 6600
     ccaactcttt ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt 6660
     ctagtgtage cgtagttagg ceaecactte aagaaetetg tageaeegee taeataeete 6720
     gctctgctaa tcctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg 6780
25 ttggactcaa gacgatagtt accggataag gcgcagcggt cgggctgaac ggggggttcg 6840 tgcacacagc ccagcttgga gcgaacgacc tacaccgaac tgagatacct acagcgtgag 6900
     ctatgagaaa gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc ggtaagcggc 6960
     agggtcggaa caggagagcg cacgagggag cttccagggg gaaacgcctg gtatctttat 7020
     agtcctgtcg ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg 7080
30
     gggcggagcc tatggaaaaa cgccagcaac gcggcctttt tacggttcct ggccttttgc 7140
     tggccttttg ctcacatgtt ctttcctgcg ttatcccctg attctgtgga taaccgtatt 7200
     accgcctttg agtgagctga taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca 7260
     gtgagcgagg aagcggaaga gcgcccaata cgcaaaccgc ctctccccgc gcgttggccg 7320
     attcattaat gcagctggca cgacaggttt cccgactgga aagcgggcag tgagcgcaac 7380
35
     gcaattaatg tgagttagct cactcattag gcaccccagg ctttacactt tatgcttccg 7440
     gctcgtatgt tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac 7500
     catgattacg ccaagctggc gcg
40
     <210> 80
     <211> 4754
     <212> DNA
     <213> Artificial Sequence
   <220>
     <223> Description of Artificial Sequence: vector
           pRK55pbsC31-Re
     <400> 80
50
     ggccgcccga tatgacacaa ggggttgtga ccggggtgga cacgtacgcg ggtgcttacg 60
     accetcagtc gegegagege gagaattega gegeageaag cecagegaca cagegtageg 120
     ccaacgaaga caaggeggcc gacettcage gegaagtega gegegaeggg ggeeggttea 180
     ggttegtegg geattteage gaagegeegg geaegtegge gttegggaeg geggagegee 240
     cggagttcga acgcatcctg aacgaatgcc gcgccgggcg gctcaacatg atcattgtct 300
55
     atgacqtqtc gcgcttctcg cgcctgaagg tcatggacgc gattccgatt gtctcggaat 360
     tgctcgccct gggcgtgacg attgtttcca ctcaggaagg cgtcttccgg cagggaaacg 420
     teatggacet gatteacetg attatgegge tegacgegte geacaaagaa tettegetga 480
     agtcggcgaa gattctcgac acgaagaacc ttcagcgcga attgggcggg tacgtcggcg 540
     ggaaggegee ttaeggette gagettgttt eggagaegaa ggagateaeg egeaaeggee 600
60
     gaatggtcaa tgtcgtcatc aacaagcttg cgcactcgac cactcccctt accggaccct 660
     tegagttega gecegaegta ateeggtggt ggtggegtga gateaagaeg cacaaacace 720
     ttcccttcaa gccgggcagt caagccgcca ttcacccggg cagcatcacg gggctttgta 780
     agegeatgga egetgaegee gtgeegaeee ggggegagae gattgggaag aagaeegett 840
     caagegeetg ggacceggea accettatge gaateetteg ggaccegegt attgeggget 900
65
     tegeegetga ggtgatetae aagaagaage eggaeggeae geegaeeaeg aagattgagg 960
     gttaccgcat tcagcgcgac ccgatcacgc tccggccggt cgagcttgat tgcggaccga 1020
     tcatcgagcc cgctgagtgg tatgagcttc aggcgtggtt ggacggcagg gggcgcggca 1080
```

PCT/EP01/12975

| | | | • | 13 | | | |
|------------|------------|------------|-------------|------------|------------|--------------|-------|
| | aggggctttc | ccgggggcaa | gccattctgt | ccqccatqqa | caagctgtac | tacaaatata | 1140 |
| | gcgccgtcat | gacttcgaag | cacaaaaaaa | aatcoatcaa | ggactettae | cactaccata | 1200 |
| | accadaadat | dutcuscada | tecacaceta | aacogaccaa | ggacccccac | age egecyce | 1000 |
| | +~~~~~ | ggtcgacccg | | ggcagcacga | aggcacgtgc | aacgtcagca | 1260 |
| 5 | rggcggcact | cgacaagttc | gttgcggaac | gcatcttcaa | caagatcagg | cacgccgaag | 1320 |
| 3 | gcgacgaaga | gacgttggcg | cttctqtqqq | aagccgcccg | acgettegge | aageteactg | 1380 |
| | aggcgcctga | gaagagcggc | gaacgggcga | accttottoc | adadededec | gacgccctga | 1440 |
| | acoccettoa | agagctgtac | ussusceace | Caaaaaaaaa | ggagegee | gacgeeeega | 1500 |
| | ddaadcactt | ccadaadaa | gaagaccgcg | tagaaggaga | gcacgacgga | ceegergeea | 1300 |
| | ggaagcaccc | ccggaagcaa | caggeagege | tgacgctccg | gcagcaaggg | gcggaagagc | 1560 |
| 10 | ggcttgccga | acttgaagcc | gccgaagccc | cgaagcttcc | ccttgaccaa | tggttccccg | 1620 |
| 10 | aagacgccga | cgctgacccg | accggcccta | agtcgtggtg | aaaacacaca | tcagtagacg | 1680 |
| | acaagcgcgt | gttcgtcggg | ctcttcgtag | acaagatcgt | tatcacaaaa | tcaactacaa | 1740 |
| | gcagggggca | gggaacgccc | atchanaanc | acacttcast | caccacgaag | 22222222 | 1000 |
| | ccascascas | 9992209000 | accgagaage | gegeeeegae | cacguagaga | aagccgccga | 1800 |
| | ccgacgacga | cgaagacgac | geecaygaeg | gcacggaaga | cgtagcggcg | taggcggcgc | 1860 |
| 1.5 | ccgggctcga | gggggggccc | ggtacccagc | ttttgttccc | tttagtgagg | gttaatttcg | 1920 |
| 15 | agcttggcgt | aatcatggtc | atagctgttt | cctqtqtqaa | attottatcc | gctcacaatt | 1980 |
| | ccacacaaca | tacgageegg | aagcataaag | totaaageet | agagtaceta | atgagtgagg | 2040 |
| | taactcacat | taattgcgtt | acactcacta | accepte | 9999090000 | acquigage | 2010 |
| | carctacatt | 2244224 | gegeteacty | CCCGCCCCC | agicgggaaa | congregage | 2100 |
| | tagetgeatt | aatgaatcgg | ccaacgcgcg | gggagaggcg | gtttgcgtat | tgggcgctct | 2160 |
| 20 | teegetteet | cgctcactga | ctcgctgcgc | teggtegtte | ggctgcggcg | agcggtatca | 2220 |
| 20 | gctcactcaa | aggcggtaat | acggttatcc | acagaatcag | gggataacgc | aggaaagaac | 2280 |
| | atgtgagcaa | aaggccagca | aaaggccagg | aaccotaaaa | aggeegett | actaccattt | 2340 |
| | ttccataggc | teegeeeee | tracracrat | Cacaaaaata | aggeegegee | tosas | 2340 |
| | Cassaccas | coogcocccc | - gacgagcat | Cacaaaaacc | gacgereaag | ccagaggtgg | 2400 |
| | tataateega | caggactata | aagataccag | gegttteece | ctggaagctc | cctcgtgcgc | 2460 |
| 25 | rerectiffe | cgaccctgcc | gcttaccgga | tacctgtccg | cetttetee | ttcgggaagc | 2520 |
| 23 | gtggcgcttt | ctcatagete | acgctqtagg | tatctcagtt | caatataaat | catteactec | 2580 |
| | aagctgggct | gtgtgcacga | accccccctt | cagcccgacc | actacacett | atcongtaac | 2640 |
| | tatcotctto | agtccaaccc | ggtaagacac | gacttatege | cactageage | accoggiage | 2700 |
| | aacammatta | acsasacasa | atatataaa | gacccaccgc | | agccactygt | 2700 |
| | 22242222 | gcagagcgag | grargragge | ggrgeraeag | agttettgaa | gragrageer | 2/60 |
| 30 | aactacggct | acactagaag | gacagtattt | ggtatctgcg | ctctgctgaa | gccagttacc | 2820 |
| 30 | ttcggaaaaa | gagttggtag | ctcttgatcc | ggcaaacaaa | ccaccactaa | tagcagtagt | 2880 |
| | ttttttgttt | gcaagcagca | gattacgcgc | agaaaaaag | gateteaaga | agateetttg | 2940 |
| | atcttttcta | cggggtctga | cactcaataa | aacraaaact | cacattaaga | asttttaata | 3000 |
| | atmamattat | caaaaaaaat | ottonestas | adegadadee | cacgicaagg | gactitiggto | 2000 |
| | tanatatana | caaaaaggat | cttcacctag | accoutttaa | attaaaaatg | aagttttaaa | 3060 |
| 35 | ccaacccaaa | gtatatatga | gtaaacttgg | tctgacagtt | accaatgctt | aatcagtgag | 3120 |
| 22 | gcacctatct | cagcgatctg | tctatttcgt | tcatccatag | ttgcctgact | ccccgtcgtg | 3180 |
| | tagataacta | cgatacggga | gggcttacca | tctqqcccca | gtgctgcaat | gataccccca | 3240 |
| | gacccacgct | caccggctcc | agatttatca | gcaataaacc | auccauccuu | 3=aaaccaaa | 3300 |
| | cacagaagta | gtcctgcaac | tttatcccc | tocatocact | atattaatta | ttaagggccgag | 3360 |
| | agtagaageg | gtttttt | | Localdeage | Clattaattg | ttgccgggaa | 3360 |
| 40 | gccayaycaa | gtagttcgcc | agttaatagt | ttgcgcaacg | ttgttgccat | tgctacaggc | 3420 |
| 40 | atcgtggtgt | cacgctcgtc | gtttggtatg | gcttcattca | gctccggttc | ccaacgatca | 3480 |
| | aggcgagtta | catgatecee | catgttgtgc | aaaaaagcgg | ttagctcctt | caatcctcca | 3540 |
| | atcgttgtca | gaagtaagtt | gaccacaata | ttatcactca | taattataac | agractorat | 3600 |
| | aattetetta | ctgtcatgcc | atcontagna | tacttttcta | tasetaatas | atacteres | 3660 |
| | aagtcattot | asasstsate | tetessea | cyclicity | Lyactygtya | gracicaacc | 2000 |
| 45 | aagccacccc | gagaatagtg | Laigeggega | ccgagttgct | cttgcccggc | gtcaatacgg | 3720 |
| 73 | gataataccg | cgccacatag | cagaacttta | aaagtgctca | tcattggaaa | acgttcttcg | 3780 |
| | gggcgaaaac | tctcaaggat | cttaccgctg | ttgagatcca | gttcgatgta | acccactcot | 3840 |
| | gcacccaact | gatcttcagc | atcttttact | ttcaccageg | fttctaaata | aucaaaaaca | 3900 |
| | ggaaggcaaa | atgccgcaaa | assonnasta | 2000000000 | aaaaatatta | ageadadaca | 3060 |
| | ctcttccttt | ttasstatta | ttassasst | t-t | ggaaacgccg | aatactcata | 3900 |
| 50 | | ttcaatatta | Ligaagcall | tatcagggtt | attgtctcat | gagcggatac | 4020 |
| 50 | atatttgaat | gtatttagaa | aaataaacaa | ataggggttc | cgcgcacatt | tccccgaaaa | 4080 |
| | gtgccaccta | aattgtaagc | gttaatattt | tgttaaaatt | cgcgttaaat | ttttgttaaa | 4140 |
| | tcagctcatt | ttttaaccaa | taggccgaaa | teggeaaaat | cccttataaa | tcaaaanaat | 4200 |
| | agaccgagat | agggttgagt | attattaaa | +++ | coccacaaa | ttaaaagaac | 40.00 |
| | tagactoco | -systemay. | gregereay | tatata | gagtecacta | ccaaagaacg | 4200 |
| 55 | tggactccaa | cyccaaaggg | cyaaaaaccg | LCEATCAGGG | cgatggccca | ctacgtgaac | 4320 |
|)) | catcacccta | atcaagtttt | ttggggtcga | ggtgccgtaa | agcactaaat | cggaacccta | 4380 |
| | aagggagccc | ccgatttaga | gcttgacggg | gaaaqccqqc | gaacgtggcg | agaaaggaag | 4440 |
| | ggaagaaagc | gaaaggagcg | gacactagga | Cactuacasa | tataacaata | acact acaca | 4500 |
| | taaccaccac | acccgccgcg | ottaatooo | cactacac | -yeageggee | torestte | 4500 |
| | aactaccac | accogacage | | cyclacaggg | egegreceat | regecattea | 4560 |
| 4 0 | ggurgcgcaa | ctgttgggaa | gggcgatcgg | rgcgggcctc | ttcgctatta | cgccagctgg | 4620 |
| 60 | cgaaaggggg | atgtgctgca | aggcgattaa | gttgggtaac | accadaattt | tcccagtcac | 4680 |
| | gacgttgtaa | aacqacqqcc | agtgaattqt | aatacqactc | actataggg | gaattagaage | 4740 |
| | tccaccgcgg | tage | | 3 | | J | 4754 |
| | | J J - | | | | | |
| | | | | | | | |

<210> 81 <211> 4773 <212> DNA 65

```
acgtacgcgg gtgcttacga ccgtcagtcg cgcgagcgcg agaattcgag cgcagcaagc 120
10
     ccagcgacac agcgtagcgc caacgaagac aaggcggccg accttcagcg cgaagtcgag 180
     cgcgacgggg gccggttcag gttcgtcggg catttcagcg aagcgccggg cacgtcggcg 240
     ttcgggacgg cggagcgccc ggagttcgaa cgcatcctga acgaatgccg cgccgggcgg 300
     ctcaacatga tcattgtcta tgacgtgtcg cgcttctcgc gcctgaaggt catggacgcg 360
     atteegattg teteggaatt getegeeetg ggegtgaega ttgttteeae teaggaagge 420
15
     gtcttccggc agggaaacgt catggacctg attcacctga ttatgcggct cgacgcgtcg 480
     cacaaagaat cttcgctgaa gtcggcgaag attctcgaca cgaagaacct tcagcgcgaa 540
     ttgggcgggt acgtcggcgg gaaggcgcct tacggcttcg agcttgtttc ggagacgaag 600
     gagatcacge gcaacggccg aatggtcaat gtcgtcatca acaagcttgc gcactcgacc 660
     acteceetta eeggaceett egagttegag eeegacgtaa teeggtggtg gtggegtgag 720
20
     atcaagacgc acaaacacct tcccttcaag ccgggcagtc aagccgccat tcacccgggc 780
     agcatcacgg ggctttgtaa gcgcatggac gctgacgccg tgccgacccg gggcgagacg 840
     attgggaaga agaccgcttc aagcgcctgg gacccggcaa ccgttatgcg aatccttcgg 900
     gaccegegta ttgcgggett egecgetgag gtgatetaca agaagaagee ggacggcacg 960 eegaccacga agattgaggg ttacegeatt cagegegace egateaeget eeggeeggte 1020
25
     gagettgatt geggacegat categagece getgagtggt atgagettea ggegtggttg 1080
     gacggcaggg ggcgcggcaa ggggctttcc cggggggcaag ccattctgtc cgccatggac 1140
     aagctgtact gcgagtgtgg cgccgtcatg acttcgaagc gcggggaaga atcgatcaag 1200
     gactettace getgeegteg eeggaaggtg gtegaceegt eegcacetgg geageacgaa 1260
     ggcacgtgca acgtcagcat ggcggcactc gacaagttcg ttgcggaacg catcttcaac 1320
     aagatcaggc acgccgaagg cgacgaagag acgttggcgc ttctgtggga agccgcccga 1380
     cgcttcggca agctcactga ggcgcctgag aagagcggcg aacgggcgaa ccttgttgcg 1440
     gagegegeg aegecetgaa egecettgaa gagetgtaeg aagacegege ggeaggegeg 1500
     tacgacggac ccgttggcag gaagcacttc cggaagcaac aggcagcgct gacgctccgg 1560
     cagcaagggg cggaagagcg gcttgccgaa cttgaagccg ccgaagcccc gaagcttccc 1620
35
     cttgaccaat ggttccccga agacgccgac gctgacccga ccggccctaa gtcgtggtgg 1680
     gggcgcgcgt cagtagacga caagcgcgtg ttcgtcgggc tcttcgtaga caagatcgtt 1740
     gtcacgaagt cgactacggg cagggggcag ggaacgccca tcgagaagcg cgcttcgatc 1800
     acgtgggcga agccgccgac cgacgacgac gaagacgacg cccaggacgg cacggaagac 1860
     gtageggegt aggeggegee egggetegag ggggggeeeg gtaceeaget tttgtteeet 1920
40
     ttagtgaggg ttaatttcga gcttggcgta atcatggtca tagctgtttc ctgtgtgaaa 1980
     ttgttatccg ctcacaattc cacacaacat acgagccgga agcataaagt gtaaagcctg 2040
     gggtgcctaa tgagtgagct aactcacatt aattgcgttg cgctcactgc ccgctttcca 2100
     gtcgggaaac ctgtcgtgcc agctgcatta atgaatcggc caacgcgcgg ggagaggcgg 2160
     tttgcgtatt gggcgctctt ccgcttcctc gctcactgac tcgctgcgct cggtcgttcg 2220
45
     gctgcggcga gcggtatcag ctcactcaaa ggcggtaata cggttatcca cagaatcagg 2280
     ggataacgca ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa 2340
     ggccgcgttg ctggcgtttt tccataggct ccgccccct gacgagcatc acaaaaatcg 2400
     acgeteaagt cagaggtgge gaaaceegae aggaetataa agataceagg egttteecee 2460
     tggaagetee etegtgeget etectgttee gaeeetgeeg ettaceggat acetgteege 2520
50
     ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggt atctcagttc 2580
     ggtgtaggtc gttcgctcca agctgggctg tgtgcacgaa ccccccgttc agcccgaccg 2640
     ctgcgcctta tccggtaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc 2700
     actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 2760
     gttcttgaag tggtggccta actacggcta cactagaagg acagtatttg gtatctgcgc 2820
55
     tetgetgaag ceagttacet teggaaaaag agttggtage tettgateeg geaaacaaac 2880
     caccgctggt agcggtggtt tttttgtttg caagcagcag attacgcgca gaaaaaaagg 2940
     atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtgga acgaaaactc 3000
     acgttaaggg attttggtca tgagattatc aaaaaggatc ttcacctaga tccttttaaa 3060
    . ttaaaaaatga agttttaaat caatctaaag tatatatgag taaacttggt ctgacagtta 3120
60
     ccaatgctta atcagtgagg cacctatctc agcgatctgt ctatttcgtt catccatagt 3180
     tgcctgactc cccgtcgtgt agataactac gatacgggag ggcttaccat ctggccccag 3240
     tgctgcaatg ataccgcgag acccacgctc accggctcca gatttatcag caataaacca 3300
     gccagccgga agggccgagc gcagaagtgg tcctgcaact ttatccgcct ccatccagtc 3360
     tattaattgt tgccgggaag ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt 3420
65
     tgttgccatt gctacaggca tcgtggtgtc acgctcgtcg tttggtatgg cttcattcag 3480
     ctccggttcc caacgatcaa ggcgagttac atgatccccc atgttgtgca aaaaagcggt 3540
     tageteette ggteeteega tegttgteag aagtaagttg geegeagtgt tateaeteat 3600
```

```
ggttatggca gcactgcata attotottac tgtcatgcca tccgtaagat gcttttctgt 3660
     gactggtgag tactcaacca agtcattctg agaatagtgt atgcggcgac cgagttgctc 3720
     ttgcccggcg tcaatacggg ataataccgc gccacatagc agaactttaa aagtgctcat 3780 cattggaaaa cgttcttcgg ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag 3840
     ttcgatgtaa cccactcgtg cacccaactg atcttcagca tcttttactt tcaccagcgt 3900
     ttctgggtga gcaaaaacag gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg 3960
     gaaatgttga atactcatac tcttcctttt tcaatattat tgaagcattt atcagggtta 4020
     ttgtctcatg agcggataca tatttgaatg tatttagaaa aataaacaaa taggggttcc 4080
     gcgcacattt ccccgaaaag tgccacctaa attgtaagcg ttaatatttt gttaaaattc 4140 gcgttaaatt tttgttaaat cagctcattt tttaaccaat aggccgaaat cggcaaaatc 4200
10
     ccttataaat caaaagaata gaccgagata gggttgagtg ttgttccagt ttggaacaag 4260
     agtecactat taaagaacgt ggactecaac gteaaaggge gaaaaaccgt etateaggge 4320
     gatggcccac tacgtgaacc atcaccctaa tcaagttttt tggggtcgag gtgccgtaaa 4380
     gcactaaatc ggaaccctaa agggagcccc cgatttagag cttgacgggg aaagccggcg 4440
15
     aacgtggcga gaaaggaagg gaagaaagcg aaaggagcgg gcgctagggc gctggcaagt 4500
     gtageggtea egetgegegt aaccaccaca ecegeegege ttaatgegee getacaggge 4560
     gcgtcccatt cgccattcag gctgcgcaac tgttgggaag ggcgatcggt gcgggcctct 4620
     tcgctattac gccagctggc gaaaggggga tgtgctgcaa ggcgattaag ttgggtaacg 4680
     ccagggtttt cccagtcacg acgttgtaaa acgacggcca gtgaattgta atacgactca 4740
20
     ctatagggcg aattggagct ccaccgcggt ggc
                                                                             4773
     <210> 82
     <211> 7803
25
     <212> DNA
     <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: vector
30
            pPGKattA1
     <400> 82
     tatcatgtct ggatccgcgt taacacctaa gaaggcgaag ttttccttac accttgcaga 60
     tataaagtgt ctaacagttt aaaatatccg atgaggcata tttatgttgg acccgtagct 120
     cagccaggat agagcactgg cctccggagc cggaggtccc gggttcaaat cccggcgggt 180
     ccgtatatta ctttttgatt cagattagat ttgtaaatct ttattacaag gataatttga 240
     tcttgtatat tggtaactct ctactctata atttttatga gaaattcaca gtcgtccctt 300
     tataccataa atagetaagt ttgtcaaagt tettattaaa etetecatgt agagattaaa 360
     teggateeat aacttegtat ageataeatt ataegaagtt ataeegggee aceatggteg 420
40
     cgagtagett ggeactggee gtegttttae aacgtegtga etgggaaaac cetggegtta 480
     cccaacttaa tcgccttgca gcacatcccc ctttcgccag ctggcgtaat agcgaagagg 540
     cccgcaccga tcgcccttcc caacagttgc gcagcctgaa tggcgaatgg cgctttgcct 600
     ggtttccggc accagaagcg gtgccggaaa gctggctgga gtgcgatctt cctgaggccg 660
     atactgtcgt cgtcccctca aactggcaga tgcacggtta cgatgcgccc atctacacca 720
45
     acgtaaccta teccattacg gteaateege egtttgttee caeggagaat eegaegggtt 780
     gttactcgct cacatttaat gttgatgaaa gctggctaca ggaaggccag acgcgaatta 840
     tttttgatgg cgttaactcg gcgtttcatc tgtggtgcaa cgggcgctgg gtcggttacg 900
     gccaggacag tcgtttgccg tctgaatttg acctgagcgc atttttacgc gccggagaaa 960
     accgcctcgc ggtgatggtg ctgcgttgga gtgacggcag ttatctggaa gatcaggata 1020
tgtggcggat gagcggcatt ttccgtgacg tctcgttgct gcataaaccg actacacaaa 1080
50
     tcagcgattt ccatgttgcc actcgcttta atgatgattt cagccgcgct gtactggagg 1140
     ctgaagttca gatgtgcggc gagttgcgtg actacctacg ggtaacagtt tctttatggc 1200
     agggtgaaac gcaggtcgcc agcggcaccg cgcctttcgg cggtgaaatt atcgatgagc 1260
     gtggtggtta tgccgatcgc gtcacactac gtctgaacgt cgaaaacccg aaactgtgga 1320
55
     gcgccgaaat cccgaatctc tatcgtgcgg tggttgaact gcacaccgcc gacggcacgc 1380
     tgattgaagc agaagcctgc gatgtcggtt tccgcgaggt gcggattgaa aatggtctgc 1440 tgctgctgaa cggcaagccg ttgctgattc gaggcgttaa ccgtcacgag catcatcctc 1500
     tgcatggtca ggtcatggat gagcagacga tggtgcagga tatcctgctg atgaagcaga 1560
     acaactttaa cgccgtgcgc tgttcgcatt atccgaacca tccgctgtgg tacacgctgt 1620
60
     gcgaccgcta cggcctgtat gtggtggatg aagccaatat tgaaacccac ggcatggtgc 1680
     caatgaatcg totgaccgat gatccgcgct ggctaccggc gatgagcgaa cgcgtaacgc 1740
     gaatggtgca gcgcgatcgt aatcacccga gtgtgatcat ctggtcgctg gggaatgaat 1800
     caggccacgg cgctaatcac gacgcgctgt atcgctggat caaatctgtc gatccttccc 1860
     gcccggtgca gtatgaaggc ggcggagccg acaccacggc caccgatatt atttgcccga 1920
65
     tgtacgcgcg cgtggatgaa gaccagccct tcccggctgt gccgaaatgg tccatcaaaa 1980
     aatggettte getacetgga gagaegegee egetgateet ttgegaatae geecaegega 2040
     tgggtaacag tcttggcggt ttcgctaaat actggcaggc gtttcgtcag tatccccgtt 2100
```

```
tacagggcgg cttcgtctgg gactgggtgg atcagtcgct gattaaatat gatgaaaacg 2160
     gcaacccgtg gtcggcttac ggcggtgatt ttggcgatac gccgaacgat cgccagttct 2220
     gtatgaacgg tctggtcttt gccgaccgca cgccgcatcc agcgctgacg gaagcaaaac 2280
     accagcagca gtttttccag ttccgtttat ccgggcaaac catcgaagtg accagcgaat 2340
     acctgttccg tcatagcgat aacgagctcc tgcactggat ggtggcgctg gatggtaagc 2400
     cgctggcaag cggtgaagtg cctctggatg tcgctccaca aggtaaacag ttgattgaac 2460
     tgcctgaact accgcagccg gagagcgccg ggcaactctg gctcacagta cgcgtagtgc 2520
     aaccgaacgc gaccgcatgg tcagaagccg ggcacatcag cgcctggcag cagtggcgtc 2580
     tggcggaaaa cctcagtgtg acgctccccg ccgcgtccca cgccatcccg catctgacca 2640
10
     ccagcgaaat ggatttttgc atcgagctgg gtaataagcg ttggcaattt aaccgccagt 2700
     caggetteet tecacagatg tggattggcg ataaaaaaca actgetgacg eegetgegeg 2760
     atcagttcac ccgtgcaccg ctggataacg acattggcgt aagtgaagcg acccgcattg 2820
     accetaaege etgggtegaa egetggaagg eggegggeea ttaccaggee gaageagegt 2880
     tgttgcagtg cacggcagat acacttgctg atgcggtgct gattacgacc gctcacgcgt 2940
     ggcagcatca ggggaaaacc ttatttatca gccggaaaac ctaccggatt gatggtagtg 3000
     gtcaaatggc gattaccgtt gatgttgaag tggcgagcga tacaccgcat ccggcgcgga 3060
     ttggcctgaa ctgccagctg gcgcaggtag cagagcgggt aaactggctc ggattagggc 3120
     cgcaagaaaa ctatcccgac cgccttactg ccgcctgttt tgaccgctgg gatctgccat 3180
     tgtcagacat gtataccccg tacgtcttcc cgagcgaaaa cggtctgcgc tgcgggacgc 3240
20
     gcgaattgaa ttatggccca caccagtggc gcggcgactt ccagttcaac atcagccgct 3300
     acagteaaca geaactgatg gaaaccagee ategecatet getgeacgeg gaagaaggea 3360
     catggctgaa tatcgacggt ttccatatgg ggattggtgg cgacgactcc tggagcccgt 3420
     cagtategge ggaatteeag etgagegeeg gtegetaeea ttaceagttg gtetggtgte 3480 aaaaataata ataaceggge aggggggate tttgtgaagg aacettaett etgtggtgtg 3540
     acataattgg acaaactacc tacagagatt taaagctcta aggtaaatat aaaattttta 3600
     agtgtataat gtgttaaact actgattcta attgtttgtg tattttagat tccaacctat 3660
     ggaactgatg aatgggagca gtggtggaat gccagatcca gacatgataa gatacattga 3720
     tgagtttgga caaaccacaa ctagaatgca gtgaaaaaaa tgctttattt gtgaaatttg 3780
     tgatgctatt gctttatttg taaccattat aagctgcaat aaacaagtta acaacaacaa 3840
30
     ttgcattcat tttatgtttc aggttcaggg ggaggtgtgg gaggtttttt aaagcaagta 3900
     aaacctctac aaatgtggta tggctgatta tgatctgcgg ccgcagggcc tcgtgatacg 3960
     cctattttta taggttaatg tcatgataat aatggtttct tagacgtcag gtggcacttt 4020
     tcggggaaat gtgcgcggaa cccctatttg tttattttc taaatacatt caaatatgta 4080
     teegeteatg agacaataac cetgataaat getteaataa tattgaaaaa ggaagagtat 4140
     gagtattcaa catttccgtg tcgcccttat tccctttttt gcggcatttt gccttcctgt 4200
     ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt tgggtgcacg 4260
     agtgggttac atcgaactgg atctcaacag cggtaagatc cttgagagtt ttcgccccga 4320
     agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgcgg tattatcccg 4380
     tattgacgcc gggcaagagc aactcggtcg ccgcatacac tattctcaga atgacttggt 4440
     tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg 4500
     cagtgctgcc ataaccatga gtgataacac tgcggccaac ttacttctga caacgatcgg 4560
     aggaccgaag gagctaaccg cttttttgca caacatgggg gatcatgtaa ctcgccttga 4620
     tcgttgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca ccacgatgcc 4680
     tgtagcaatg gcaacaacgt tgcgcaaact attaactggc gaactactta ctctagcttc 4740
45
     ccggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac ttctgcgctc 4800
     ggcccttccg gctggctggt ttattgctga taaatctgga gccggtgagc gtgggtctcg 4860
     cggtatcatt gcagcactgg ggccagatgg taagccctcc cgtatcgtag ttatctacac 4920
     gacggggagt caggcaacta tggatgaacg aaatagacag atcgctgaga taggtgcctc 4980
     actgattaag cattggtaac tgtcagacca agtttactca tatatacttt agattgattt 5040
50
     aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata atctcatgac 5100
     caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag aaaagatcaa 5160
     aggatettet tgagateett tttttetgeg egtaatetge tgettgeaaa caaaaaaace 5220
     accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt ttccgaaggt 5280
     aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc cgtagttagg 5340
55
     ccaccacttc aagaactctg tagcaccgcc tacatacctc gctctgctaa tcctgttacc 5400
     agtggctgct gccagtggcg ataagtcgtg tcttaccggg ttggactcaa gacgatagtt 5460
     accggataag gcgcagcggt cgggctgaac ggggggttcg tgcacacagc ccagcttgga 5520 gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa gcgccacgct 5580
     tcccgaaggg agaaaggcgg acaggtatcc ggtaagcggc agggtcggaa caggagagcg 5640
60
     cacgagggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg ggtttcgcca 5700
     cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc tatggaaaaa 5760
     cgccagcaac gcggcctttt tacggttcct ggccttttgc tggccttttg ctcacatgtt 5820
     ctttcctgcg ttatcccctg attctgtgga taaccgtatt accgcctttg agtgagctga 5880
     taccgetege egeageegaa egacegageg cagegagtea gtgagegagg aageggaaga 5940
65
     gegeceaata egeaaacege eteteceege gegttggeeg atteattaat geagetggea 6000
     cgacaggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg tgagttagct 6060
     cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt tgtgtggaat 6120
```

```
tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg ccaagctggc 6180
     gcgtcgagga attctaccgg gtaggggagg cgcttttccc aaggcagtct ggagcatgcg 6240
     ctttagcage eccgetggge acttggeget acacaagtgg ectetggete geacacatte 6300
     cacatccacc ggtaggcgcc aaccggctcc gttctttggt ggccccttcg cgccaccttc 6360
     tactcctccc ctagtcagga agttcccccc cgccccgcag ctcgcgtcgt gcaggacgtg 6420
     acaaatggaa gtagcacgtc tcactagtct cgtgcagatg gacagcaccg ctgagcaatg 6480
     gaagegggta ggeetttggg geageggeea atageagett tgeteetteg etttetggge 6540
     gggcgcccga aggtcctccg gaggcccggc attctgcacg cttcaaaagc gcacgtctgc 6660
10
     egegetgtte teetetteet eateteeggg cetttegace tgeageeegg tacagttega 6720
     ataacttcgt atagcataca ttatacgaag ttataagctt gcatgcctgc aggtcggccg 6780
     ccacgacegg ccggccggtg ccgccaccat cccctgaccc acgcccctga cccctcacaa 6840
     ggagacgacc ttccatgacc gagtacaagc ccacggtgcg cctcgccacc cgcgacgacg 6900
     tecceeggge egtacgeace etegeegeeg egttegeega etaceeegee aegegeeaca 6960
15
     ccgtcgaccc ggaccgccac atcgagcggg tcaccgagct gcaagaactc ttcctcacgc 7020
     gcgtcgggct cgacatcggc aaggtgtggg tcgcggacga cggcgccgcg gtggcggtct 7080
     ggaccacgcc ggagagcgtc gaagcggggg cggtgttcgc cgagatcggc ccgcgcatgg 7140
     ccgagttgag cggttcccgg ctggccgcgc agcaacagat ggaaggcctc ctggcgccgc 7200
     accggcccaa ggagcccgcg tggttcctgg ccaccgtcgg cgtctcgccc gaccaccagg 7260
20
     gcaagggtet gggcagcgcc gtcgtgctcc ccggagtgga ggcggccgag cgcgccgggg 7320
     tcaccgtcac cgccgacgtc gagtgcccga aggaccgcgc gacctggtgc atgacccgca 7440
     ageceggtge etgaegeeeg eeceaegaee egeagegeee gaeegaaagg agegeaegae 7500
     cccatggete cgaccgaage cgacccggge ggecccgccg accccgcace cgccccgag 7560
25
     gcccaccgac tctagaggat cataatcagc cataccacat ttgtagaggt tttacttgct 7620
     ttaaaaaacc tcccacacct ccccctgaac ctgaaacata aaatgaatgc aattgttgtt 7680
     gttaacttgt ttattgcagc ttataatggt tacaaataaa gcaatagcat cacaaatttc 7740
     acaaataaag cattttttc actgcattct agttgtggtt tgtccaaact catcaatgta 7800
30
     <210> 83
     <211> 8167
     <212> DNA
35
     <213> Artificial Sequence
     <220>
    <223> Description of Artificial Sequence: vector
         pPGKattA2
40
     <400> 83 ·
     tatcatgtct ggatccgcgt taacacctaa gaaggcgaag ttttccttac accttgcaga 60
     tataaagtgt ctaacagttt aaaatatccg atgaggcata tttatgttgg acccgtagct 120
     cagccaggat agagcactgg ceteeggage eggaggtece gggtteaaat eeeggegggt 180
45
     cogtatatta ctttttgatt cagattagat ttgtaaatct ttattacaag gataatttga 240
     tettgtatat tggtaactet etactetata atttttatga gaaatteaca gtegteeett 300
     tataccataa atagctaagt ttgtcaaagt tcttattaaa ctctccatgt agagattaaa 360
     teggatecat aacttegtat ageatacatt atacgaagtt atacegggee accatggteg 420
     cgagtagett ggcactggcc gtcgttttac aacgtcgtga ctgggaaaac cctggcgtta 480
50
     cccaacttaa tegeettgea geacateeee etttegeeag etggegtaat agegaagagg 540
     cccgcaccga tcgcccttcc caacagttgc gcagcctgaa tggcgaatgg cgctttgcct 600
     ggtttccggc accagaagcg gtgccggaaa gctggctgga gtgcgatctt cctgaggccg 660
     atactgtcgt cgtcccctca aactggcaga tgcacggtta cgatgcgccc atctacacca 720
     acgtaaccta teccattaeg gteaateege egtttgttee caeggagaat eegaegggtt 780
55
     gttactcgct cacatttaat gttgatgaaa gctggctaca ggaaggccag acgcgaatta 840
     tttttgatgg cgttaactcg gcgtttcatc tgtggtgcaa cgggcgctgg gtcggttacg 900
     gccaggacag tcgtttgccg tctgaatttg acctgagcgc atttttacgc gccggagaaa 960
     accgcctcgc ggtgatggtg ctgcgttgga gtgacggcag ttatctggaa gatcaggata 1020
     tgtggcggat gagcggcatt ttccgtgacg tctcgttgct gcataaaccg actacacaaa 1080
60
     tcagcgattt ccatgttgcc actcgcttta atgatgattt cagccgcgct gtactggagg 1140
     ctgaagttca gatgtgcggc gagttgcgtg actacctacg ggtaacagtt tctttatggc 1200 agggtgaaac gcaggtcgcc agcggcaccg cgcctttcgg cggtgaaatt atcgatgagc 1260
     gtggtggtta tgccgatcgc gtcacactac gtctgaacgt cgaaaacccg aaactgtgga 1320
     gegeegaaat eccgaatete tategtgegg tggttgaaet geacacegee gaeggeaege 1380
65
     tgattgaagc agaagcctgc gatgtcggtt tccgcgaggt gcggattgaa aatggtctgc 1440
     tgctgctgaa cggcaagccg ttgctgattc gaggcgttaa ccgtcacgag catcatcctc 1500
     tgcatggtca ggtcatggat gagcagacga tggtgcagga tatcctgctg atgaagcaga 1560
```

```
acaactttaa cgccgtgcgc tgttcgcatt atccgaacca tccgctgtgg tacacgctgt 1620
     gcgaccgcta cggcctgtat gtggtggatg aagccaatat tgaaacccac ggcatggtgc 1680
     caatgaatcg totgaccgat gatccgcgct ggctaccggc gatgagcgaa cgcgtaacgc 1740
     gaatggtgca gcgcgatcgt aatcacccga gtgtgatcat ctggtcgctg gggaatgaat 1800
     caggecaegg egetaateae gaegegetgt ategetggat caaatetgte gateetteee 1860
     geceggtgea gtatgaagge ggeggageeg acaceaegge cacegatatt atttgeeega 1920
     tgtacgcgcg cgtggatgaa gaccagccct tcccggctgt gccgaaatgg tccatcaaaa 1980
     aatggettte getaeetgga gagaegegee egetgateet ttgegaatae geceaegega 2040
     tgggtaacag tettggeggt ttegetaaat actggeagge gtttegteag tateceegtt 2100
10
     tacagggcgg cttcgtctgg gactgggtgg atcagtcgct gattaaatat gatgaaaacg 2160 gcaacccgtg gtcggcttac ggcggtgatt ttggcgatac gccgaacgat cgccagttct 2220
     gtatgaacgg tctggtcttt gccgaccgca cgccgcatcc agcgctgacg gaagcaaaac 2280
     accagcagca gtttttccag ttccgtttat ccgggcaaac catcgaagtg accagcgaat 2340
     acctgttccg tcatagcgat aacgagctcc tgcactggat ggtggcgctg gatggtaagc 2400
15
     cgctggcaag cggtgaagtg cctctggatg tcgctccaca aggtaaacag ttgattgaac 2460
     tgcctgaact accgcagccg gagagcgccg ggcaactctg gctcacagta cgcgtagtgc 2520
     aaccgaacge gaccgcatgg teagaageeg ggcacateag egeetggcag cagtggegte 2580
     tggcggaaaa cctcagtgtg acgctccccg ccgcgtccca cgccatcccg catctgacca 2640
     ccagcgaaat ggatttttgc atcgagctgg gtaataagcg ttggcaattt aaccgccagt 2700
20
     caggetteet tteacagatg tggattggcg ataaaaaaca actgetgacg ecgetgegeg 2760
     atcagttcac ccgtgcaccg ctggataacg acattggcgt aagtgaagcg acccgcattg 2820
     accetaacge etgggtegaa egetggaagg eggegggeea ttaccaggee gaageagegt 2880
     tgttgcagtg cacggcagat acacttgctg atgcggtgct gattacgacc gctcacgcgt 2940
     ggcagcatca ggggaaaacc ttatttatca gccggaaaac ctaccggatt gatggtagtg 3000
25
     gtcaaatggc gattaccgtt gatgttgaag tggcgagcga tacaccgcat ccggcgcgga 3060
     ttggcctgaa ctgccagctg gcgcaggtag cagagcgggt aaactggctc ggattagggc 3120
     cgcaagaaaa ctatcccgac cgccttactg ccgcctgttt tgaccgctgg gatctgccat 3180
     tgtcagacat gtataccccg tacgtcttcc cgagcgaaaa cggtctgcgc tgcgggacgc 3240
     gcgaattgaa ttatggccca caccagtggc gcggcgactt ccagttcaac atcagccgct 3300
30
     acagtcaaca gcaactgatg gaaaccagcc atcgccatct gctgcacgcg gaagaaggca 3360
     catggctgaa tatcgacggt ttccatatgg ggattggtgg cgacgactcc tggagcccgt 3420
     cagtategge ggaatteeag etgagegeeg gtegetacea ttaceagttg gtetggtgte 3480 aaaaataata ataaceggge aggggggate tttgtgaagg aacettactt etgtggtgtg 3540
     acataattgg acaaactacc tacagagatt taaagctcta aggtaaatat aaaattttta 3600
35
     agtgtataat gtgttaaact actgattcta attgtttgtg tattttagat tccaacctat 3660
     ggaactgatg aatgggagca gtggtggaat gccagatcca gacatgataa gatacattga 3720
     tgagtttgga caaaccacaa ctagaatgca gtgaaaaaaa tgctttattt gtgaaatttg 3780
     tgatgctatt gctttatttg taaccattat aagctgcaat aaacaagtta acaacaacaa 3840
     ttgcattcat tttatgtttc aggttcaggg ggaggtgtgg gaggtttttt aaagcaagta 3900 aaacctctac aaatgtggta tggctgatta tgatctgcgg ccgcagggcc tcgtgatacg 3960
40
     cctattttta taggttaatg tcatgataat aatggtttct tagacgtcag gtggcacttt 4020
     tcggggaaat gtgcgcggaa cccctatttg tttatttttc taaatacatt caaatatgta 4080
     teegeteatg agacaataae eetgataaat getteaataa tattgaaaaa ggaagagtat 4140
     gagtattcaa cattteegtg tegecettat teeetttttt geggeatttt geetteetgt 4200 ttttgeteac ccagaaacge tggtgaaagt aaaagatget gaagateagt tgggtgeacg 4260
45
     agtgggttac atcgaactgg atctcaacag cggtaagatc cttgagagtt ttcgccccga 4320
     agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgcgg tattatcccg 4380
     tattgacgcc gggcaagagc aactcggtcg ccgcatacac tattctcaga atgacttggt 4440
     tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg 4500
50
     cagtgctgcc ataaccatga gtgataacac tgcggccaac ttacttctga caacgatcgg 4560
     aggaccgaag gagctaaccg cttttttgca caacatgggg gatcatgtaa ctcgccttga 4620
     tegttgggaa eeggagetga atgaageeat accaaaegae gagegtgaea eeaegatgee 4680
     tgtagcaatg gcaacaacgt tgcgcaaact attaactggc gaactactta ctctagcttc 4740
     ccggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac ttctgcgctc 4800
55
     ggcccttccg gctggctggt ttattgctga taaatctgga gccggtgagc gtgggtctcg 4860
     cggtatcatt gcagcactgg ggccagatgg taagccctcc cgtatcgtag ttatctacac 4920 gacggggagt caggcaacta tggatgaacg aaatagacag atcgctgaga taggtgcctc 4980
     actgattaag cattggtaac tgtcagacca agtttactca tatatactit agattgattt 5040
     aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata atctcatgac 5100
60
     caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag aaaagatcaa 5160
     aggatettet tgagateett tttttetgeg egtaatetge tgettgeaaa caaaaaaace 5220
     accyctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt ttccgaaggt 5280
     aactggette agcagagege agataceaaa tactgteett etagtgtage egtagttagg 5340
     ccaccacttc aagaactetg tagcaccgcc tacatacctc getetgetaa teetgttacc 5400
65
     agtggctgct gccagtggcg ataagtcgtg tcttaccggg ttggactcaa gacgatagtt 5460
     accggataag gcgcagcggt cgggctgaac ggggggttcg tgcacacagc ccagcttgga 5520
     gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa gcgccacgct 5580
```

```
tecegaaggg agaaaggegg acaggtatee ggtaagegge agggteggaa caggagageg 5640
     cacgagggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg ggtttcgcca 5700
     cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc tatggaaaaa 5760
     cgccagcaac gcggcctttt tacggttcct ggccttttgc tggccttttg ctcacatgtt 5820
     cttteetgeg ttateecetg attetgtgga taaccgtatt accgeetttg agtgagetga 5880
     taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg aagcggaaga 5940
     gegeceaata egeaaacege eteteceege gegttggeeg atteattaat geagetggea 6000
     cgacaggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg tgagttagct 6060
     cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt tgtgtggaat 6120
     tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg ccaagctggc 6180
     gcgtcgagga attctaccgg gtaggggagg cgcttttccc aaggcagtct ggagcatgcg 6240
     ctttagcagc cccgctgggc acttggcgct acacaagtgg cctctggctc gcacacattc 6300
     cacatceace ggtaggegec aaceggetee gttetttggt ggeeeetteg egecacette 6360
     tactectece etagteagga agttececce egeecegeag etegegtegt geaggacgtg 6420
15
     acaaatggaa gtagcacgtc tcactagtct cgtgcagatg gacagcaccg ctgagcaatg 6480
     gaagcgggta ggcctttggg gcagcggcca atagcagctt tgctccttcg ctttctgggc 6540
     gggcgcccga aggtcctccg gaggcccggc attctgcacg cttcaaaagc gcacgtctgc 6660
     cgcgctgttc tcctcttcct catctccggg cctttcgacc tgcagcccgg tacagttcga 6720
     aggateegeg ttaacaceta agaaggegaa gtttteetta cacettgeag atataaagtg 6780
     tctaacagtt taaaatatcc gatgaggcat atttatgttg gacccgtagc tcagccagga 6840
     tagagcactg gcctccggag ccggaggtcc cgggttcaaa tcccggcggg tccgtatatt 6900
     actttttgat tcagattaga tttgtaaatc tttattacaa ggataatttg atcttgtata 6960
     ttggtaactc tctactctat aatttttatg agaaattcac agtcgtccct ttataccata 7020
. 25
     aatagctaag tttgtcaaag ttcttattaa actctccatg tagagattaa atcggatcct 7080
     tcgaataact tcgtatagca tacattatac gaagttataa gcttgcatgc ctgcaggtcg 7140
     geogecacga coggeoggee ggtgeogeca coatecoetg acceaegece etgacecete 7200
     acaaggagac gaccttccat gaccgagtac aagcccacgg tgcgcctcgc cacccgcgac 7260
     gacgtccccc gggccgtacg caccctcgcc gccgcgttcg ccgactaccc cgccacgcgc 7320
30
     cacaccgtcg acccggaccg ccacatcgag cgggtcaccg agctgcaaga actcttcctc 7380
     acgcgcgtcg ggctcgacat cggcaaggtg tgggtcgcgg acgacggcgc cgcggtggcg 7440
     gtctggacca cgccggagag cgtcgaagcg ggggcggtgt tcgccgagat cggcccgcgc 7500
     atggccgagt tgagcggttc ccggctggcc gcgcagcaac agatggaagg cctcctggcg 7560
     ccgcaccggc ccaaggagcc cgcgtggttc ctggccaccg tcggcgtctc gcccgaccac 7620
     cagggcaagg gtctgggcag cgccgtcgtg ctccccggag tggaggcggc cgagcgcgcc 7680 ggggtgcccg ccttctgga gacctccgcg ccccgcaacc tccccttcta cgagcggctc 7740
     ggetteaceg teacegeega egtegagtge eegaaggaee gegegaeetg gtgeatgaee 7800
     cgcaagcccg gtgcctgacg cccgcccac gacccgcagc gcccgaccga aaggagcgca 7860
     cgacccatg gctccgaccg aagccgaccc ggggggcccc gccgaccccg cacccgcccc 7920
     cgaggcccac cgactctaga ggatcataat cagccatacc acatttgtag aggttttact 7980
     tgctttaaaa aacctcccac acctcccct gaacctgaaa cataaaatga atgcaattgt 8040
     tgttgttaac ttgtttattg cagcttataa tggttacaaa taaagcaata gcatcacaaa 8100
     tttcacaaat aaagcatttt tttcactgca ttctagttgt ggtttgtcca aactcatcaa 8160
     tgtatct
45
     <210> 84
     <211> 51
     <212> DNA
50
     <213> Artificial Sequence
     <223> Description of Artificial Sequence: primer XisAl
55
     <400> 84
     ataagaatgc ggccgcccga tatgcaaaat cagggtcaag acaaatatca a
                                                                        51
     <210> 85
60
     <211> 76
     <212> DNA
     <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: primer XisA2
     <400> 85
```

```
ataagaatgc ggccgcacca tgcccaagaa gaagaggaag gtgcaaaatc agggtcaaga 60
     caaatatcaa caagec
 5
     <210> 86
     <211> 44
     <212> DNA
     <213> Artificial Sequence
10
     <220>
     <223> Description of Artificial Sequence: primer XisA3
     <400> 86
     ataagaatgc ggccgctcaa ctattcttat aagctatttc catc
                                                                         44
15
     <210> 87
     <211> 82
     <212> DNA
20
     <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: primer nifD1
25
     <400> 87
     cgatggctct tcccttccgt caaatgcact cttgggatta ctccgaacct agcgatgggg 60
     tgcaaatgtc agatcagata ag
30
     <210> 88
     <211> 82
     <212> DNA
     <213> Artificial Sequence
35
     <223> Description of Artificial Sequence: primer nifD2
     <400> 88
     cgcttatctg atctgacatt tgcaccccat cgctaggttc ggagtaatcc caagagtgca 60
40
     tttgacggaa gggaagagcc at
     <210> 89
     <211> 74
45
     <212> DNA
     <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: primer nifD3
50
     <400> 89
     gatcagctgt tgaaagctat taaaccacaa aaaggattac tccggccctt atcacggtta 60
     cgacggattt gcta
55
     <210> 90
     <211> 74
     <212> DNA
     <213> Artificial Sequence
60
     <220>
     <223> Description of Artificial Sequence: primer nifD4
65
     gatctagcaa atccgtcgta accgtgataa gggccggagt aatccttttt gtggtttaat 60
     agctttcaac agct
```

```
<210> 91
     <211> 41
     <212> DNA
     <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: primerSSV1-1
10
    <400> 91
     ataagaatgc ggccgcccga tatgacgaaa gataagacgc g
                                                                        41
     <210> 92
15
     <211> 70
     <212> DNA
     <213> Artificial Sequence
     <220>
20
     <223> Description of Artificial Sequence: primer SSV1-2
     <400> 92
     ataagaatgc ggccgcacca tgcccaagaa gaagaggaag gtgacgaaag ataagacgcg 60
     ttataaatac
25
     <210> 93
     <211> 47
     <212> DNA
30
     <213> Artificial Sequence
     <220>
    <223> Description of Artificial Sequence: primer SSV2
35
     <400> 93
     tgtcccgggc tcgaaaccgg ggggatccgc ttgtagggga gtatccc
                                                                        47
     <210> 94
40
     <211> 47
     <212> DNA
     <213> Artificial Sequence
     <220>
45
     <223> Description of Artificial Sequence: primer SSV3
     <400> 94
     gagcccggga caagcggaag cggtggtgga aaagagggaa ctgaacg
                                                                        47
50
     <210> 95
     <211> 33
     <212> DNA
     <213> Artificial Sequence
55
     <220>
     <223> Description of Artificial Sequence: primer SSV4
     <400> 95
60
     atcgctcgag tcagacccct tttagccatt ccg
                                                                        33
     <210> 96
     <211> 40
65
     <212> DNA
     <213> Artificial Sequence
```

tacgacggac ccgttggcag gaagcacttc cggaagcaac aggcagcgct gacgctccgg 1560

```
cagcaagggg cggaagagcg gettgccgaa ettgaagccg ccgaagcccc gaagcttccc 1620
     cttgaccaat ggttccccga agacgccgac gctgacccga ccggccctaa gtcgtggtgg 1680
     gggcgcgcgt cagtagacga caagcgcgtg ttcgtcgggc tcttcgtaga caagatcgtt 1740
     gtcacgaagt cgactacggg cagggggcag ggaacgccca tcgagaagcg cgcttcgatc 1800
     acgtgggcga agccgccgac cgacgacgac gaagacgacg cccaggacgg cacggaagac 1860
     gtageggegt ageggeeget etagaactag tggateeece gggetgeagg aattegatat 1920
     caagettate gatacegteg acetegaggg ggggeeeggt acceagettt tgtteeettt 1980
     agtgagggtt aatttcgagc ttggcgtaat catggtcata gctgtttcct gtgtgaaatt 2040
     gttatecget cacaatteca cacaacatae gageeggaag cataaagtgt aaageetggg 2100
10
     cgggaaacct gtcgtgccag ctgcattaat gaatcggcca acgcgcgggg agaggcggtt 2220
     tgcgtattgg gcgctcttcc gcttcctcgc tcactgactc gctgcgctcg gtcgttcggc 2280
     tgcggcgagc ggtatcagct cactcaaagg cggtaatacg gttatccaca gaatcagggg 2340
     ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa ggccaggaac cgtaaaaagg 2400
15
     ccgcgttgct ggcgtttttc cataggctcc gccccctga cgagcatcac aaaaatcgac 2460
     getcaagtca gaggtggcga aacccgacag gactataaag ataccaggcg tttcccctg 2520
     gaageteect egtgegetet cetgtteega eeetgeeget taeeggatae etgteegeet 2580
     ttetecette gggaagegtg gegetttete atageteaeg etgtaggtat eteagttegg 2640
     tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag cccgaccgct 2700
20
     gcgccttatc cggtaactat cgtcttgagt ccaacccggt aagacacgac ttatcgccac 2760
     tggcagcagc cactggtaac aggattagca gagcgaggta tgtaggcggt gctacagagt 2820
     tettgaagtg gtggcctaac tacggctaca ctagaaggac agtatttggt atctgcgctc 2880
     tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc aaacaaacca 2940
     ccgctggtag cggtggtttt tttgtttgca agcagcagat tacgcgcaga aaaaaaggat 3000
25
     ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc tcagtggaac gaaaactcac 3060
     gttaagggat tttggtcatg agattatcaa aaaggatctt cacctagatc cttttaaatt 3120
     aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggtct gacagttacc 3180
     aatgettaat cagtgaggea cetateteag egatetgtet atttegttea tecatagttg 3240
     cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct ggccccagtg 3300
30
     ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca ataaaccagc 3360
     cageeggaag ggeegagege agaagtggte etgeaacttt ateegeetee ateeagteta 3420
     ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtttg cgcaacgttg 3480
     ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggtatggct tcattcagct 3540
     ccggttccca acgatcaagg cgagttacat gatcccccat gttgtgcaaa aaagcggtta 3600
35
     getecttegg tecteegate gttgteagaa gtaagttgge egeagtgtta teacteatgg 3660
     ttatggcage actgcataat tetettaetg teatgecate egtaagatge ttttetgtga 3720
     ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg agttgctctt 3780 gcccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa gtgctcatca 3840
     ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt 3900
40
     cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc accagcgttt 3960
     ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg gcgacacgga 4020
     aatgttgaat actcatactc ttcctttttc aatattattg aagcatttat cagggttatt 4080
     gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata ggggttccgc 4140
     gcacatttee eegaaaagtg eeacetaaat tgtaagegtt aatattttgt taaaattege 4200
45
     gttaaatttt tgttaaatca gctcattttt taaccaatag gccgaaatcg gcaaaatccc 4260
     ttataaatca aaagaataga ccgagatagg gttgagtgtt gttccagttt ggaacaagag 4320
     tccactatta aagaacgtgg actccaacgt caaagggcga aaaaccgtct atcagggcga 4380
     tggcccacta cgtgaaccat caccctaatc aagttttttg gggtcgaggt gccgtaaagc 4440
     actaaatcgg aaccctaaag ggagcccccg atttagagct tgacggggaa agccggcgaa 4500
50
     cgtggcgaga aaggaaggga agaaagcgaa aggagcgggc gctagggcgc tggcaagtgt 4560
     ageggteaeg etgegegtaa ceaecaeaee egeegegett aatgegeege taeagggege 4620
     gtcccattcg ccattcaggc tgcgcaactg ttgggaaggg cgatcggtgc gggcctcttc 4680
     gctattacgc cagctggcga aagggggatg tgctgcaagg cgattaagtt gggtaacgcc 4740
     agggttttcc cagtcacgac gttgtaaaac gacggccagt gaattgtaat acgactcact 4800
55
     atagggcgaa ttggagctcc accgcggtgg c
     <210> 100
     <211> 20
60
     <212> DNA
     <213> Artificial Sequence
     <220>
     <223> Description of Artificial Sequence: primer
65
          C31-screen 1
     <400> 100
```

| | | | | 0.5 | | | |
|----|-------------|------------------|-------------|-------------------|------------|-------------|-------|
| | gccatacact | tgagtgacat | tgacatccac | tttgcctttc | totocacago | tatccactcc | 960 |
| | Caggggggggg | tcgaccatgc | CCaanaanaa | andandta. | tocaatttao | taracataa | 1020 |
| | Connette | cotacoutgo | coadgaagaa | gaggaaggcg | | tyaccytaca | 1020 |
| | ccaaaattty | cctgcattac | eggregatge | aacgagtgat | gaggttcgca | agaacctgat | 1080 |
| _ | ggacatgttc | agggatcgcc | aggcgttttc | tgagcatacc | tggaaaatgc | ttctqtccqt | 1140 |
| 5 | ttgccggtcg | tgggcggcat | ggtgcaagtt | gaataaccgg | aaatggtttc | ссосадаасс | 1200 |
| | tgaagatgtt | cgcgattatc | ttotatatot | †03000000 | aatataaaa | ***** | 1000 |
| | concernent | tteeseesee | teecacaccc | ccaggegege | ggcccggcag | Ladaaactat | 1200 |
| | Coagoacat | ttgggccagc | taaacatgct | tcatcgtcgg | tccgggctgc | cacgaccaag | 1320 |
| | tgacagcaat | gctgtttcac | tggttatgcg | gcggatccga | aaagaaaacg | ttgatgccgg | 1380 |
| | tgaacgtgca | aaacaggctc | tagcgttcga | acqcactgat | ttcgaccagg | ttcattcact | 1440 |
| 10 | catogaaaat | agcgatcgct | accaddatat | acataatata | acatttataa | aasttaatts | 1500 |
| | taaggaaaata | ttaartataa | gocaggatat | acycaaccig | gcarricing | ggattyctta | 1300 |
| | Laacaccccg | ttacgtatag | ccgaaattgc | caggatcagg | gttaaagata | tctcacgtac | 1560 |
| | tgacggtggg | agaatgttaa | tccatattgg | cagaacgaaa | acgctggtta | gcaccgcagg | 1620 |
| | tgtagagaag | gcacttagcc | tgggggtaac | taaactggtc | gagcgatgga | tttccatctc | 1680 |
| | tagtataget | gatgatccga | ataactacct | attttaccaa | atasassas | atastattas | 1740 |
| 15 | Oggegeaget | ga c ga c c c ga | 222222222 | grangergg | gicayaaaaa | arggrated | 1/40 |
| 10 | cycyccatct | gccaccagcc | agctatcaac | regegeeetg | gaagggattt | ttgaagcaac | T800: |
| | tcatcgattg | atttacggcg | ctaaggatga | ctctggtcag | agatacctgg | cctggtctgg | 1860 |
| | acacagtgcc | cgtgtcggag | ccqcqcqaqa | tataacccac | actagaattt | caataccoga | 1920 |
| | gatcatgcaa | gctggtggct | ggaccaatgt | aaatattoto | atraactata | tocataacat | 1080 |
| | nnatantnaa | 344334334 | taataaaaat | ~~t~~~~i~t | acquactata | | 1300 |
| 20 | - sacagegaa | acaggggcaa | eggegegeee | gerggaagar | ggcgattage | cattaacgcg | 2040 |
| 20 | taaatgattg | cagatccact | agttctaggg | ccgcgtcgac | ctcgagatcc | aggcgcggat | 2100 |
| | caataaaaga | tcattatttt | caatagatct | gtgtgttggt | tttttqtqtq | ccttggggga | 2160 |
| | gagagagacc | agaatgaggc | gcggccaagg | audadddda | uuccauaatu | accttaggg | 2220 |
| | 22222222 | agaaataaaa | ttaaaaaaa | 999499994 | ggccagaacg | accetggggg | 2220 |
| | agggggaggc | cagaatgacc | rrgggggagg | gggaggccag | aargaggcgc | gcccccgggt | 2280 |
| 25 | accgageteg | aattcactgg | ccgtcgtttt | acaacgtcgt | gactgggaaa | accctggcgt | 2340 |
| 25 | tacccaactt | aatcgccttg | cagcacatcc | ccctttcgcc | agctggcgta | atagcgaaga | 2400 |
| | ggcccgcacc | gatcgccctt | cccaacagtt | acacaaccta | aatuucuaat | ancacctast | 2460 |
| | acaatattt | ctccttacgc | atctatages | +++++ | aacggcgaac | ggcgcccgac | 2500 |
| | t | tetectacyc | accegegegg | Latticacac | cgcatatggt | gcactctcag | 2520 |
| | Lacaatetge | tctgatgccg | catagttaag | ccagccccga | cacccgccaa | cacccgctga | 2580 |
| | cgcgccctga | cgggcttgtc | tgctcccggc | atccgcttac | agacaagctg | tgaccgtctc | 2640 |
| 30 | cgggagctgc | atototcaga | ggttttcacc | gtcatcaccg | aaacgcgcga | nnnasanna | 2700 |
| | cctcgtgata | cacctatttt | tatadettaa | tatastasts | 2+22+00++ | attagaaaggg | 2760 |
| | 2000gcgaca | ttt | -t-t | tyttatyata | acaacygeee | cctagacgte | 2/00 |
| | aggtggcact | tttcggggaa | argracacaa | aacccctatt | tgtttattt | tctaaataca | 2820 |
| | ttcaaatatg | tatccgctca | tgagacaata | accctgataa | atgcttcaat | aatattgaaa | 2880 |
| | aaggaagagt | atgagtattc | aacatttccg | tatcaccctt | attecetttt | ttgcggcatt | 2940 |
| 35 | ttgccttcct | attitacte | acccanaaac | actaatass | ntaaaanatn | ctassastas | 3000 |
| | attacataca | | acctagaaac | geeggegaaa | gradadary | ctyaayatta | 3000 |
| | gttgggtgca | cgagrgggrt | acategaact | ggatctcaac | agcggtaaga | tccttgagag | 3060 |
| | ttttcgcccc | gaagaacgtt | ttccaatgat | gagcactttt | aaagttctgc | tatgtggcgc | 3120 |
| | ggtattatcc | cgtattgacg | ccqqqcaaqa | gcaactcggt | coccocatac | actattctca | 3180 |
| | gaatgacttg | attaaatact | caccactcac | anaaaancat | cttacccatc | acataacaat | 3340 |
| 40 | 3222222 | tacatacta | oaccagccac | agaaaagcac | cccacggacg | gcatgatagt | 2240 |
| 70 | aayayaatta | tgcagtgctg | ccataaccat | gagtgataac | actgcggcca | acttacttct | 3300 |
| , | gacaacgatc | ggaggaccga | aggagctaac | cgcttttttg | cacaacatgg | gggatcatgt | 3360 |
| | aactcgcctt | gatcgttggg | aaccggaget | gaatgaagcc | ataccaaacq | acqaqcqtqa | 3420 |
| | caccaccato | cctgtagcaa | toocaacaac | attacacasa | ctattaacto | accesactact | 3400 |
| | tactotaccet | tocatagoaa | 22442242 | gt cgcgcaaa | ccaccaaccy | ycyaactact | 3400 |
| 45 | acticiaget | teceggeaac | aactaga | crygatggag | geggataaag | LIGCAGGACC | 3540 |
| 43 | acttetgege | taggacatta | cggctggctg | gtttattgct | gataaatctg | gagccggtga | 3600 |
| | gcgtgggtct | cgcggtatca | ttgcagcact | ggggccagat | ggtaagccct | cccatatcat | 3660 |
| | agttatetae | acgacgggga | gtcaggcaac | tatogatosa | cgaaatagac | agategetga | 3720 |
| | gataggtgcc | tcactgatta | agoattggta | 2019500500 | anathtach | | 3700 |
| | ttacattact | ttanantt | abbett | tana | Lagittact | Lacatatatt | 3700 |
| 50 | ctagattgat | ttaaaacttc | atttttaatt | taaaaggatc | taggtgaaga | tcctttttga | 3840 |
| 50 | taatctcatg | accaaaatcc | cttaacgtga | gttttcgttc | cactgagcgt | cagaccccgt | 3900 |
| | agaaaagatc | aaaggatctt | cttgagatcc | tttttttcta | cocotaatct | actacttaca | 3960 |
| | Aacaaaaaa | ccaccgctac | caggggtggt | ttatttaaaa | antannasa | tagazzatat | 4020 |
| | *** | -t | cageggegge | ctgcttgccg | gattaagagt | Laccaactet | 4020 |
| | | gtaactggct | | | | | |
| | gccgtagtta | ggccaccact | tcaagaactc | tgtagcaccg | cctacatacc | tegetetget | 4140 |
| 55 | aatcctgtta | ccagtggctg | ctaccaataa | cgataagtcg | tatettacea | ggttggactc | 4200 |
| | aanacratar | ttaccggata | 20000000000 | ataggees | 20000000 | 990099000 | 4260 |
| | augucgatag | ccaccygaca | aggegeageg | gccgggccga | acggggggtt | cytycacaca | 4200 |
| | gcccagcttg | gagcgaacga | cctacaccga | actgagatac | ctacagcgtg | agctatgaga | 4320 |
| | aagcgccacg | cttcccgaag | ggagaaaggc | ggacaggtat | ccqqtaaqcq | gcagggtcgg | 4380 |
| | aacaggagag | cgcacgaggg | agettecagn | gggaaacgcc | tagtatettt | atagteetet | 4440 |
| 60 | Cuuct++-~~ | Cacctotos | ttasaaata | 3554446466C | hash | acageocege | 4E00 |
| UU | - aggettege | cacctctgac | Ligagogiog | allitgtga | rgcrcgrcag | gggggcggag | 4500 |
| | CCTatggaaa | aacgccagca | acgcggcctt | tttacggttc | ctggcctttt | gctggccttt | 4560 |
| | tgctcacatq | ttctttcctg | cgttatcccc | tgattctgtg | gataaccota | ttaccocctt | 4620 |
| | tgagtgagcf | gataccgctc | accacaaacca | aacdaccdad | Cucsucusus | cautuaucua | 4690 |
| | adaz | gacaccac | taggagagag | accyactyay | ogcaycyayt | cagigagiga | 4040 |
| 65 | ggaagcggaa | gagegeedaa | Lacycaaacc | yeardteece | gcgcgttggc | cgattcatta | 4/40 |
| 65 | atgcagctgg | cacgacaggt | ttcccgactg | gaaagcgggc | agtgagcgca | acgcaattaa | 4800 |
| | tgtgagttag | ctcactcatt | aggcacccca | ggctttacac | tttatocttc | cggctcgtat | 4860 |
| | gttgtgtgga | attgtgagcg | gataacaatt | tcacacanna | aacadctato | accatgatta | 4920 |
| | - 5-5-55 | | J | uggu | -acayceacy | | |
| | | | | | | | |

```
cgccaagcta gcccgggcta gcttgcatgc ctgcaggttt tcgacattga ttattgacta 4980
     gttattaata gtaatcaatt acggggtcat tagttcatag cccatatatg gagttccgcg 5040
     ttacataact tacggtaaat ggcccgcctg gctgaccgcc caacgacccc cgcccattga 5100
     cgtcaataat gacgtatgtt cccatagtaa cgccaatagg gactttccat tgacgtcaat 5160
     gggtggacta tttacggtaa actgcccact tggcagtaca tcaagtgtat catatgccaa 5220
     gtacgccccc tattgacgtc aatgacggta aatggcccgc ctggcattat gcccagtaca 5280
     tgaccttatg ggactttcct acttggcagt acatctacgt attagtcatc gctattacca 5340
     tgggtcgagg tgagccccac gttctgcttc actctcccca tctcccccc ctccccaccc 5400
     ccaattttgt atttattat tttttaatta ttttgtgcag cgatgggggc ggggggggg 5460
10
     ggggcgcgc ccaggcgggg cggggcggg cgaggggcgg ggcggggggg ggcggagagg 5520
     tgcggcggca gccaatcaga gcggcgcgct ccgaaagttt ccttttatgg cgaggcggcg 5580
     gcggcggcgg ccctataaaa agcgaagcgc gcggcgggcg ggagtcgctg cgttgccttc 5640
     geccegtgce cegeteegeg cegeetegeg cegeeegeee eggetetgae tgacegegtt 5700 acteceacag gtgageggge gggacegee tteteeteeg ggetgtaatt agegettggt 5760
15
     ttaatgacgg ctcgtttctt ttctgtggct gcgtgaaagc cttaaagggc tccgggaggg 5820
     ccctttgtgc gggggggagc ggctcggggg gtgcgtgcgt gtgtgtgtgc gtggggag
     <210> 105
     <211> 6641
     <212> DNA
     <213> Artificial Sequence
     <220>
25 <223> Description of Artificial Sequence: vector
           pCAGGSC31CNLS-pA
     <400> 105
     attgattatt gactagttat taatagtaat caattacggg gtcattagtt catagcccat 60
30
     atatggagtt ccgcgttaca taacttacgg taaatggccc gcctggctga ccgcccaacg 120
     acccccgccc attgacgtca ataatgacgt atgttcccat agtaacgcca atagggactt 180
     tccattgacg tcaatgggtg gactatttac ggtaaactgc ccacttggca gtacatcaag 240
     tgtatcatat gccaagtacg ccccctattg acgtcaatga cggtaaatgg cccgcctggc 300
     attatgccca gtacatgacc ttatgggact ttcctacttg gcagtacatc tacgtattag 360
    tcatcgctat taccatgggt cgaggtgagc cccacgttct gcttcactct ccccatctcc 420
     ccccccccc cacccccaat tttgtattta tttatttttt aattattttg tgcagcgatg 480
     ggggcgggg ggggggggc gcgcgccagg cggggcgggg cggggcgagg ggcgggggg 540
     ggcgaggcgg agaggtgcgg cggcagccaa tcagagcggc gcgctccgaa agtttccttt 600
     tatggcgagg cggcggcggcgccta taaaaagcga agcgcgcggc gggcgggagt 660
     cgctgcgttg ccttcgcccc gtgccccgct ccgcgccgcc tcgcgccgcc cgccccggct 720
     ctgactgacc gcgttactcc cacaggtgag cgggcgggac ggcccttctc ctccgggctg 780
     taattagcgc ttggtttaat gacggctcgt ttcttttctg tggctgcgtg aaagccttaa 840
     agggeteegg gagggeeett tgtgeggggg ggageggete ggggggtgeg tgegtgtgtg 900
     tgtgcgtggg gagcgccgcg tgcggcccgc gctgcccggc ggctgtgagc gctgcgggcg 960
     cggcgcgggg ctttgtgcgc tccgcgtgtg cgcgagggga gcgcggccgg gggcggtgcc 1020
     ccgcggtgcg ggggggctgc gaggggaaca aaggctgcgt gcggggtgtg tgcgtggggg 1080
     ggtgagcagg gggtgtgggc gcggcggtcg ggctgtaacc ccccctgca ccccctccc 1140 cgagttgctg agcacggcc ggcttcgggt gcggggctcc gtgcggggcg tggcgcgggg 1200
     ctcgccgtgc cgggcggggg gtggcggcag gtgggggtgc cgggcggggc ggggccgcct 1260
50
     cgggccgggg agggctcggg ggaggggcgc ggcggccccg gagcgccggc ggctgtcgag 1320
     gcgcggcgag ccgcagccat tgccttttat ggtaatcgtg cgagagggcg cagggacttc 1380
     ctttgtccca aatctggcgg agccgaaatc tgggaggcgc cgccgcaccc cctctagcgg 1440
     gcgcgggcga agcggtgcgg cgccggcagg aaggaaatgg gcggggaggg ccttcgtgcg 1500
     tegeogogo geogtecot tetecatete cageeteggg getgeogoag ggggaegget 1560
55
     gccttcgggg gggacggggc agggcggggt tcggcttctg gcgtgtgacc ggcggctcta 1620
     gtaagcgttg gggtgagtac tccctctcaa aagcgggcat gacttctgcg ctaagattgt 1680
     cagtttccaa aaacgaggag gatttgatat tcacctggcc cgcggtgatg cctttgaggg 1740
     tggccgcgtc catctggtca gaaaagacaa tctttttgtt gtcaagcttg aggtgtggca 1800
     ggcttgagat ctggccatac acttgagtga cattgacatc cactttgcct ttctctccac 1860
60
     aggtgtccac teccagggcg gccgcccgat atgacacaag gggttgtgac cggggtggac 1920
     acgtacgcgg gtgcttacga ccgtcagtcg cgcgagcgcg agaattcgag cgcagcaagc 1980
     ccagcgacac agcgtagcgc caacgaagac aaggcggccg accttcagcg cgaagtcgag 2040
     cgcgacgggg gccggttcag gttcgtcggg catttcagcg aagcgccggg cacgtcggcg 2100
     ttcgggacgg cggagcgcc ggagttcgaa cgcatcctga acgaatgccg cgccgggcgg 2160
65
     ctcaacatga tcattgtcta tgacgtgtcg cgcttctcgc gcctgaaggt catggacgcg 2220
     attecgatty teteggaatt getegeeetg ggegtgaega ttgttteeae teaggaagge 2280
     gtcttccggc agggaaacgt catggacctg attcacctga ttatgcggct cgacgcgtcg 2340
```

```
cacaaagaat cttcgctgaa gtcggcgaag attctcgaca cgaagaacct tcagcgcgaa 2400
      ttgggcgggt acgtcggcgg gaaggcgcct tacggcttcg agcttgtttc ggagacgaag 2460
      gagateaege geaacggeeg aatggteaat gtegteatea acaagettge geactegace 2520
      actoccotta coggacoctt cgagttogag cocgacgtaa tooggtggtg gtggcgtgag 2580 atcaagacgo acaaacacct toocttoaag cogggcagto aagcogocat toaccogggo 2640
      agcatcacgg ggctttgtaa gcgcatggac gctgacgccg tgccgacccg gggcgagacg 2700
      attgggaaga agaccgcttc aagcgcctgg gacccggcaa ccgttatgcg aatccttcgg 2760
      gaccegegta ttgcgggctt cgccgctgag gtgatctaca agaagaagcc ggacggcacg 2820
      ccgaccacga agattgaggg ttaccgcatt cagcgcgacc cgatcacgct ccggccggtc 2880
 10
      gagettgatt geggaeegat categageee getgagtggt atgagettea ggegtggttg 2940
      gacggcaggg ggcgcggcaa ggggctttcc cgggggcaag ccattctgtc cgccatggac 3000
      aagctgtact gcgagtgtgg cgccgtcatg acttcgaagc gcggggaaga atcgatcaag 3060
     gactettace getgeegteg ceggaaggtg gtegaceegt eegeacetgg geageacgaa 3120 ggeacgtgea acgteageat ggeggeacte gacaagtteg ttgeggaacg catetteaac 3180
 15
      aagatcaggc acgccgaagg cgacgaagag acgttggcgc ttctgtggga agccgcccga 3240
      cgcttcggca agctcactga ggcgcctgag aagagcggcg aacgggcgaa ccttgttgcg 3300
      gagcgcgccg acgccctgaa cgcccttgaa gagctgtacg aagaccgcgc ggcaggcgcg 3360
      tacgacggac ccgttggcag gaagcacttc cggaagcaac aggcagcgct gacgctccgg 3420
      cagcaagggg cggaagagcg gcttgccgaa cttgaagccg ccgaagcccc gaagcttccc 3480
 20
     cttgaccaat ggttccccga agacgccgac gctgacccga ccggccctaa gtcgtggtgg 3540
      gggcgcgcgt cagtagacga caagcgcgtg ttcgtcgggc tcttcgtaga caagatcgtt 3600
     gtcacgaagt cgactacggg cagggggcag ggaacgccca tcgagaagcg cgcttcgatc 3660
     acgtgggcga agccgccgac cgacgacgac gaagacgacg cccaggacgg cacggaagac 3720
     gtagcggcgc ctaagaagaa gaggaaggtt tagactctcg agatccaggc gcggatcaat 3780
     aaaagatcat tattitcaat agatctgigt gtiggtttti tgtgtgccit gggggagggg 3840
     gaggccagaa tgaggcgcgg ccaaggggga gggggaggcc agaatgacct tgggggaggg 3900
     ggaggccaga atgaccttgg gggaggggga ggccagaatg aggcgcgccc ccgggtaccg 3960
     agctcgaatt cactggccgt cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc 4020
     caacttaatc gccttgcagc acatccccct ttcgccagct ggcgtaatag cgaagaggcc 4080
30
     cgcaccgatc gcccttccca acagttgcgc agcctgaatg gcgaatggcg cctgatgcgg 4140
     tattttetee ttacgcatet gtgeggtatt teacacegea tatggtgeae teteagtaca 4200
     atetgetetg atgeegeata gttaageeag eecegacace egecaacace egetgaegeg 4260
     ccctgacggg cttgtctgct cccggcatcc gcttacagac aagctgtgac cgtctccggg 4320
     agetgeatgt gteagaggtt tteacegtea teacegaaac gegegagaeg aaagggeete 4380
35
     gtgatacgcc tatttttata ggttaatgtc atgataataa tggtttctta gacgtcaggt 4440
     ggcacttttc ggggaaatgt gcgcggaacc cctatttgtt tatttttcta aatacattca 4500
     aatatgtate egeteatgag acaataacee tgataaatge tteaataata ttgaaaaagg 4560
     aagagtatga gtattcaaca tttccgtgtc gcccttattc ccttttttgc ggcattttgc 4620
     cttcctgttt ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga agatcagttg 4680
40
     ggtgcacgag tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagtttt 4740
     cgccccgaag aacgttttcc aatgatgagc acttttaaag ttctgctatg tggcgcggta 4800
     ttatcccgta ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat 4860
     gacttggttg agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga 4920
     gaattatgca gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca 4980
45
     acgatcggag gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact 5040
     cgccttgatc gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc 5100
     acgatgeetg tageaatgge aacaacgttg egeaaactat taactggega actacttact 5160
     ctagetteec ggcaacaatt aatagactgg atggaggegg ataaagttge aggaceaett 5220
     ctgcgctcgg cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt 5280
50
     gggtetegeg gtateatige ageactgggg ceagatggta ageceteceg tategtagtt 5340
     atctacacga cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata 5400
     ggtgcctcac tgattaagca ttggtaactg tcagaccaag tttactcata tatactttag 5460
     attgatttaa aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgataat 5520
     ctcatgacca aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa 5580
55
     aagatcaaag gatcttcttg agatcctttt tttctgcgcg taatctgctg cttgcaaaca 5640
     aaaaaaccac cgctaccagc ggtggtttgt ttgccggatc aagagctacc aactcttttt 5700
     ccgaaggtaa ctggcttcag cagagcgcag ataccaaata ctgtccttct agtgtagccg 5760
     tagttaggec accaetteaa gaaetetgta geaeegeeta eatacetege tetgetaate 5820
     ctgttaccag tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga 5880
60
     cgatagttac cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc 5940
     agettggage gaacgaceta caccgaactg agatacetac agegtgaget atgagaaage 6000
     gccacgette ecgaagggag aaaggeggac aggtateegg taageggeag ggteggaaca 6060
     ggagagegea egagggaget tecaggggga aacgeetggt atetttatag teetgteggg 6120
     tttegecace tetgaettga gegtegattt ttgtgatget egteaggggg geggageeta 6180
65
     tggaaaacg ccagcaacgc ggcctttta cggttcctgg ccttttgctg gccttttgct 6240
     cacatgttct ttcctgcgtt atcccctgat tctgtggata accgtattac cgcctttgag 6300
     tgagetgata eegetegeeg eageegaaeg aeegagegea gegagteagt gagegaggaa 6360
```

```
gcggaagagc gcccaatacg caaaccgcct ctccccgcgc gttggccgat tcattaatgc 6420
     agetggeaeg acaggtttee egaetggaaa gegggeagtg agegeaaege aattaatgtg 6480 agttagetea eteattagge acceeagget ttacaettta tgetteegge tegtatgttg 6540
     tgtggaattg tgagcggata acaatttcac acaggaaaca gctatgacca tgattacgcc 6600
     aagctagece gggetagett geatgeetge aggttttega e
      <210> 106
      <211> 11784
10
     <212> DNA
      <213> Artificial Sequence
      <220>
     <223> Description of Artificial Sequence: modified
15
            ROSA26 locus
     <400> 106
     ggcaggcect ccgagcgtgg tggagccgtt ctgtgagaca gccgggtacg agtcgtgacg 60
     ctggaagggg caagcgggtg gtgggcagga atgcggtccg ccctgcagca accggagggg 120
20
     gagggagaag ggagcggaaa agtctccacc ggacgcggcc atggctcggg ggggggggg 180
     cageggagga gegetteegg eegacgtete gtegetgatt ggettetttt eeteeegeeg 240
     tgtgtgaaaa cacaaatggc gtgttttggt tggcgtaagg cgcctgtcag ttaacggcag 300
     ccggagtgcg cagccgcgg cagcctcgct ctgcccactg ggtggggcgg gaggtaggtg 360
     25
     gagggtcagc gaaagtagct cgcgcgcgag cggccgccca ccctcccctt cctctggggg 480
     agtcgtttta cccgccgccg gccgggcctc gtcgtctgat tggctctcgg ggcccagaaa 540
     actggccctt gccattggct cgtgttcgtg caagttgagt ccatccgccg gccagcgggg 600
     gcggcgagga ggcgctccca ggttccggcc ctcccctcgg ccccgcgccg cagagtctgg 660
     ccgcgcgccc ctgcgcaacg tggcaggaag cgcgcgctgg gggcggggac gggcagtagg 720 gctgagcggc tgcggggcgg gtgcaagcac gtttccgact tgagttgcct caagaggggc 780
30
     gtgctgagcc agacctccat cgcgcactcc ggggagtgga gggaaggagc gagggctcag 840
     ttgggctgtt ttggaggcag gaagcacttg ctctcccaaa gtcgctctga gttgttatca 900
     gtaagggage tgcagtggag taggcgggag gaaggccgca cccttctccg gagggggag 960
     gggagtgttg caatacettt etgggagtte tetgetgeet eetggettet gaggacegee 1020
35
     ctgggcctgg gagaatccct tccccctctt ccctcgtgat ctgcaactcc agtctttcgc 1080
     ctaggtaacc gatatccctg caggggtgac ctgcacgtct agggcgcagt agtccagggt 1140
     ttccttgatg atgtcatact tatcctgtcc ctttttttc cacagctcgc ggttgaggac 1200
     aaactetteg eggtetttee agtacteetg eaggtgactg actgagtega egacactgea 1260
     gagacctact tcactaacaa ccggtacagt tcgtggacca gatgggtgag gtggagtacg 1320
40
     cgcccgggga gcccaagggc acgccctggc acccgcaccg cggcttcgag accgtcacga 1380
     ataacttcgt atagcataca ttatacgaag ttataagctc gatgaattct accgggtagg 1440
     ggaggcgctt ttcccaaggc agtctggagc atgcgcttta gcagccccgc tggcacttgg 1500
     cgctacacaa gtggcctctg gcctcgcaca cattccacat ccaccggtag cgccaaccgg 1560
     ctccgttctt tggtggcccc ttcgcgccac cttctactcc tcccctagtc aggaagttcc 1620
45
     ccccegeccc gcagctcgcg tcgtgcagga cgtgacaaat ggaagtagca cgtctcacta 1680
     gtctcgtgca gatggacagc accgctgagc aatggaagcg ggtaggcctt tggggcagcg 1740
     gccaatagca gctttgctcc ttcgctttct gggctcagag gctgggaagg ggtgggtccg 1800
    ggggggggt caggggggg ctcagggggg gggcgggggg gaaggtcctc ccgaggcccg 1860 gcattctcgc acgcttcaaa agcgcacgtc tgccgcgctg ttctcctctt cctcatctcc 1920
50
     gggcettteg acgatecage egecaccatg aaaaageetg aacteacege gacgtetgte 1980
     gagaagtttc tgatcgaaaa gttcgacagc gtctccgacc tgatgcagct ctcggagggc 2040
     gaagaatete gtgettteag ettegatgta ggagggegtg gatatgteet gegggtaaat 2100
     agetgegeeg atggttteta caaagategt tatgtttate ggeactttge ateggeegeg 2160
     ctcccgattc cggaagtgct tgacattggg gaattcagcg agagcctgac ctattgcatc 2220
55
     tecegeegtg cacagggtgt cacgttgcaa gacetgeetg aaacegaact geeegetgtt 2280
     ctgcagccgg tcgcggaggc catggatgcg atcgctgcgg ccgatcttag ccagacgagc 2340
     gggttcggcc cattcggacc gcaaggaatc ggtcaataca ctacatggcg tgatttcata 2400
     tgcgcgattg ctgatcccca tgtgtatcac tggcaaactg tgatggacga caccgtcagt 2460
     gcgtccgtcg cgcaggctct cgatgagctg atgctttggg ccgaggactg ccccgaagtc 2520
60
     cggcaccteg tgcacgcgga tttcggctcc aacaatgtcc tgacggacaa tggccgcata 2580
     acageggtea ttgactggag egaggegatg tteggggatt cecaataega ggtegeeaac 2640 atettettet ggaggeegtg gttggettgt atggageage agaegegeta ettegagegg 2700
     aggeateegg agettgeagg ategeegegg eteegggegt atatgeteeg cattggtett 2760
     gaccaactet atcagagett ggttgacggc aatttcgatg atgcagettg ggcgcagggt 2820
65
     cgatgcgacg caatcgtccg atccggagcc gggactgtcg ggcgtacaca aatcgcccgc 2880
     agaagcgcgg ccgtctggac cgatggctgt gtagaagtac tcgccgatag tggaaaccga 2940
     cgccccagca ctcgtccgag ggcaaaggaa tagtcgatgc agaaattgat gatctattaa 3000
```

```
acaataaaga tgtccactaa aatggaagtt tttcctgtca tactttgtta agaagggtga 3060
     gattagataa atgcctgctc tttactgaag gctctttact attgctttat gataatgttt 3180
     catagttgga tatcataatt taaacaagca aaaccaaatt aagggccagc tcattcctcc 3240
     cactcatgat ctatagatct atagatctct cgtgggatca ttgttttct cttgattccc 3300
     actttgtggt tctaagtact gtggtttcca aatgtgtcag tttcatagcc tgaagaacga 3360
     gatcagcage ctctgttcca catacactte attctcagta ttgttttgcc aagttctaat 3420
     tccatcagaa gcttcagctg ctcgactaga ggatcataat cagccatacc acatttgtag 3480
     aggttttact tgctttaaaa aacctcccac acctcccct gaacctgaaa cataaaatga 3540
10
     atgcaattgt tgttgttaac ttgtttattg cagcttataa tggttacaaa taaagcaata 3600
     gcatcacaaa titcacaaat aaagcattti tticactgca tictagttgt ggtttgtcca 3660
     aactcatcaa tgtatcttat catgtctgga teegtgteat gteggegaee ctaegeeece 3720
     aactgagaga actcaaaggt taccccagtt ggggcactac tcccgaaaac cgcttctgga 3780
     tccataactt cgtatagcat acattatacg aagttatacc gggccaccat ggtcgcgagt 3840
     agcttggcac tggccgtcgt tttacaacgt cgtgactggg aaaaccctgg cgttacccaa 3900
     cttaategee ttgcagcaca teeecettte geeagetgge gtaatagega agaggeeege 3960
     accgategee etteccaaca gttgegeage etgaatggeg aatggegett tgeetggttt 4020
     ccggcaccag aagcggtgcc ggaaagctgg ctggagtgcg atcttcctga ggccgatact 4080
     gtcgtcgtcc cctcaaactg gcagatgcac ggttacgatg cgcccatcta caccaacgta 4140
20
     acctatecea ttacggteaa teegeegttt gtteecaegg agaateegae gggttgttae 4200
    togotoacat ttaatgttga tgaaagotgg ctacaggaag gocagacgog aattattttt 4260.
     gatggcgtta actcggcgtt tcatctgtgg tgcaacgggc gctgggtcgg ttacggccag 4320
     gacagtegtt tgccgtctga atttgacctg agegeatttt tacgcgccgg agaaaaccgc 4380
     ctcgcggtga tggtgctgcg ttggagtgac ggcagttatc tggaagatca ggatatgtgg 4440
25
     eggatgageg geatttteeg tgaegteteg ttgetgeata aacegactae acaaatcage 4500
     gatttccatg ttgccactcg ctttaatgat gatttcagcc gcgctgtact ggaggctgaa 4560
     gttcagatgt gcggcgagtt gcgtgactac ctacgggtaa cagtttcttt atggcagggt 4620
     gaaacgcagg tegecagegg caeegegeet tteggeggtg aaattatega tgagegtggt 4680
     ggttatgccg atcgcgtcac actacgtctg aacgtcgaaa acccgaaact gtggagcgcc 4740
30
     gaaatcccga atctctatcg tgcggtggtt gaactgcaca ccgccgacgg cacgctgatt 4800
     gaagcagaag cetgegatgt eggttteege gaggtgegga ttgaaaatgg tetgetgetg 4860
     ctgaacggca agccgttgct gattcgaggc gttaaccgtc acgagcatca tcctctgcat 4920
     ggtcaggtca tggatgagca gacgatggtg caggatatcc tgctgatgaa gcagaacaac 4980
     tttaacgccg tgcgctgttc gcattatccg aaccatccgc tgtggtacac gctgtgcgac 5040
35
     cgctacggcc tgtatgtggt ggatgaagcc aatattgaaa cccacggcat ggtgccaatg 5100
     aatcgtctga ccgatgatcc gcgctggcta ccggcgatga gcgaacgcgt aacgcgaatg 5160
     gtgcagcgcg atcgtaatca cccgagtgtg atcatctggt cgctggggaa tgaatcaggc 5220
     cacggcgcta atcacgacgc gctgtatcgc tggatcaaat ctgtcgatcc ttcccgcccg 5280
     gtgcagtatg aaggcggcgg agccgacacc acggccaccg atattatttg cccgatgtac 5340
40
     gcgcgcgtgg atgaagacca gcccttcccg gctgtgccga aatggtccat caaaaaatgg 5400
     ctttcgctac ctggagagac gcgcccgctg atcctttgcg aatacgccca cgcgatgggt 5460
     aacagtettg geggtttege taaatactgg caggegttte gteagtatee cegtttacag 5520
     ggcggcttcg tctgggactg ggtggatcag tcgctgatta aatatgatga aaacggcaac 5580 ccgtggtcgg cttacggcgg tgattttggc gatacgccga acgatcgcca gttctgtatg 5640
45
     aacggtctgg tctttgccga ccgcacgccg catccagcgc tgacggaagc aaaacaccag 5700
     cagcagtttt tocagttccg tttatccggg caaaccatcg aagtgaccag cgaatacctg 5760
     ttccgtcata gcgataacga gctcctgcac tggatggtgg cgctggatgg taagccgctg 5820
     gcaagcggtg aagtgcctct ggatgtcgct ccacaaggta aacagttgat tgaactgcct 5880
     gaactaccgc agccggagag cgccgggcaa ctctggctca cagtacgcgt agtgcaaccg 5940
50
     aacgcgaccg catggtcaga agccgggcac atcagcgcct ggcagcagtg gcgtctggcg 6000
     gaaaacctca gtgtgacgct ccccgccgcg tcccacgcca tcccgcatct gaccaccagc 6060
     gaaatggatt tttgcatcga gctgggtaat aagcgttggc aatttaaccg ccagtcaggc 6120
     tttctttcac agatgtggat tggcgataaa aaacaactgc tgacgccgct gcgcgatcag 6180
     ttcacccgtg caccgctgga taacgacatt ggcgtaagtg aagcgacccg cattgaccct 6240
55
     aacgcctggg tcgaacgctg gaaggcggcg ggccattacc aggccgaagc agcgttgttg 6300 cagtgcacgg cagatacact tgctgatgcg gtgctgatta cgaccgctca cgcgtggcag 6360
     catcagggga aaaccttatt tatcagccgg aaaacctacc ggattgatgg tagtggtcaa 6420
     atggcgatta ccgttgatgt tgaagtggcg agcgatacac cgcatccggc gcggattggc 6480
     ctgaactgcc agctggcgca ggtagcagag cgggtaaact ggctcggatt agggccgcaa 6540
60
     gaaaactatc ccgaccgcct tactgccgcc tgttttgacc gctgggatct gccattgtca 6600
     gacatgtata ccccgtacgt cttcccgagc gaaaacggtc tgcgctgcgg gacgcgcgaa 6660
     ttgaattatg gcccacacca gtggcgcggc gacttccagt tcaacatcag ccgctacagt 6720 caacagcaac tgatggaaac cagccatcgc catctgctgc acgcggaaga aggcacatgg 6780
     ctgaatatcg acggtttcca tatggggatt ggtggcgacg actcctggag cccgtcagta 6840
65
     teggeggaat tecagetgag egeeggtege taccattace agttggtetg gtgtcaaaaa 6900
     taataataac cgggcagggg ggatctttgt gaaggaacct tacttctgtg gtgtgacata 6960
     attggacaaa ctacctacag agatttaaag ctctaaggta aatataaaat ttttaagtgt 7020
```

```
ataatgtgtt aaactactga ttctaattgt ttgtgtattt tagattccaa cctatggaac 7080
     tgatgaatgg gagcagtggt ggaatgccag atccagacat gataagatac attgatgagt 7140
     ttggacaaac cacaactaga atgcagtgaa aaaaatgctt tatttgtgaa atttgtgatg 7200
     ctattgcttt atttgtaacc attataagct gcaataaaca agttaacaac aacaattgca 7260
     ttcattttat gtttcaggtt cagggggagg tgtgggaggt tttttaaagc aagtaaaacc 7320
     tctacaaatg tggtatggct gattatgatc tgcggccaaa tcggccggcc taggcgcgcc 7380
     ggtaaccgaa gttcctatac tttctagaga ataggaactt cggaatagga acttcaagct 7440
     taagegetag eetagaagat gggegggagt ettetgggea ggettaaagg etaacetggt 7500
     gtgtgggcgt tgtcctgcag gggaattgaa caggtgtaaa attggaggga caagacttcc 7560
10
     cacagatttt cggttttgtc gggaagtttt ttaatagggg caaataagga aaatgggagg 7620
     ataggtagtc atctggggtt ttatgcagca aaactacagg ttattattgc ttgtgatccg 7680
     cctcggagta ttttccatcg aggtagatta aagacatgct cacccgagtt ttatactctc 7740 ctgcttgaga tccttactac agtatgaaat tacagtgtcg cgagttagac tatgtaagca 7800
     gaattttaat catttttaaa gagcccagta cttcatatcc atttctcccg ctccttctgc 7860
15
     agocttatca aaaggtattt tagaacactc attttagccc cattttcatt tattatactg 7920
     gcttatccaa cccctagaca gagcattggc attttccctt tcctgatctt agaagtctga 7980
     tgactcatga aaccagacag attagttaca tacaccacaa atcgaggctg tagctggggc 8040
     ctcaacactg cagttetttt ataacteett agtacacttt ttgttgatee tttgeettga 8100
     teettaattt teagtgteta teacetetee egteagtggt gtteeacatt tgggeetatt 8160
20
     ctcagtccag ggagttttac aacaatagat gtattgagaa tccaacctaa agcttaactt 8220
     tocactocca tgaatgooto totoottitt otocatttat aaactgagot attaaccatt 8280
     aatggttcca ggtggatgtc tcctccccat attacctgat gtatcttaca tattgccagg 8340
     ctgatatttt aagacattaa aaggtatatt tcattattga gccacatggt attgattact 8400
     gettactaaa attttgteat tgtacacate tgtaaaaggt ggtteetttt ggaatgeaaa 8460
25
     gttcaggtgt ttgttgtctt tcctgaccta aggtcttgtg agcttgtatt ttttctattt 8520
     aagcagtgct ttotottgga etggottgae toatggoatt ctacaegtta ttgctggtet 8580
     aaatgtgatt ttgccaagct tcttcaggac ctataatttt gcttgacttg tagccaaaca 8640
     caagtaaaat gattaagcaa caaatgtatt tgtgaagctt ggtttttagg ttgttgtgtt 8700
     gtgtgtgctt gtgctctata ataatactat ccaggggctg gagaggtggc tcggagttca 8760
30
     agagcacaga ctgctcttcc agaagtcctg agttcaattc ccagcaacca catggtggct 8820
     cacaaccatc tgtaatggga tctgatgccc tcttctggtg tgtctgaaga ccacaagtgt 8880
     attcacatta aataaataaa tootoottot tottotttt titttttta aagagaatac 8940
     tgtctccagt agaatttact gaagtaatga aatactttgt gtttgttcca atatggtagc 9000
     caataatcaa attactcttt aagcactgga aatgttacca aggaactaat ttttatttga 9060
     agtgtaactg tggacagagg agccataact gcagacttgt gggatacaga agaccaatgc 9120 agactttaat gtctttctc ttacactaag caataaagaa ataaaaattg aacttctagt 9180 atcctatttg tttaaactgc tagctttact taacttttgt gcttcatcta tacaaagctg 9240
35
     aaagctaagt ctgcagccat tactaaacat gaaagcaagt aatgataatt ttggatttca 9300
     aaaatgtagg gccagagttt agccagccag tggtggtgct tgcctttatg cctttaatcc 9360
40
     cagcactctg gaggcagaga caggcagatc tctgagtttg agcccagcct ggtctacaca 9420
     tcaagttcta tctaggatag ccaggaatac acacagaaac cctgttgggg aggggggctc 9480
     tgagatttca taaaattata attgaagcat tccctaatga gccactatgg atgtggctaa 9540
     atccgtctac ctttctgatg agatttgggt attattttt ctgtctctgc tgttggttgg 9600
     gtcttttgac actgtgggct ttctttaaag cctccttcct gccatgtggt ctcttgtttg 9660
45
     ctactaactt cccatggctt aaatggcatg gctttttgcc ttctaagggc agctgctgag 9720
     atttgcagcc tgatttccag ggtggggttg ggaaatcttt caaacactaa aattgtcctt 9780
     taatttttt tttaaaaaat gggttatata ataaacctca taaaatagtt atgaggagtg 9840
     aggtggacta atattaaatg agtccctccc ctataaaaga gctattaagg ctttttgtct 9900
     tatacttaac tttttttta aatgtggtat ctttagaacc aagggtctta gagttttagt 9960
50
     atacagaaac tgttgcatcg cttaatcaga ttttctagtt tcaaatccag agaatccaaa 10020
     ttcttcacag ccaaagtcaa attaagaatt tctgactttt aatgttaatt tgcttactgt 10080
     gaatataaaa atgatagctt ttcctgaggc agggtctcac tatgtatctc tgcctgatct 10140
     gcaacaagat atgtagacta aagttetgee tgettttgte teetgaatae taaggttaaa 10200
     atgtagtaat acttttggaa cttgcaggtc agattctttt ataggggaca cactaaggga 10260
55
     gcttgggtga tagttggtaa aatgtgtttc aagtgatgaa aacttgaatt attatcaccg 10320
     caacctactt tttaaaaaaa aaagccaggc ctgttagagc atgcttaagg gatccctagg 10380
     acttgctgag cacacaagag tagttacttg gcaggctcct ggtgagagca tatttcaaaa 10440
     aacaaggcag acaaccaaga aactacagtt aaggttacct gtctttaaac catctgcata 10500
     tacacaggga tattaaaata ttccaaataa tatttcattc aagttttccc ccatcaaatt 10560
60
     gggacatgga tttctccggt gaataggcag agttggaaac taaacaaatg ttggtttgt 10620
     gatttgtgaa attgttttca agtgatagtt aaagcccatg agatacagaa caaagctgct 10680
     atttcgaggt ctcttggttt atactcagaa gcacttcttt gggtttccct gcactatcct 10740
     gatcatgtgc taggcctacc ttaggctgat tgttgttcaa ataaacttaa gtttcctgtc 10800
     aggtgatgtc atatgatttc atatatcaag gcaaaacatg ttatatatgt taaacatitg 10860
65
     tacttaatgt gaaagttagg tetttgtggg tttgattttt aatttteaaa acctgageta 10920
     aataagtcat ttttacatgt cttacatttg gtggaattgt ataattgtgg tttgcaggca 10980
     agactetetg acctagtaac cetacetata gagcactttg etgggteaca agtetaggag 11040
```

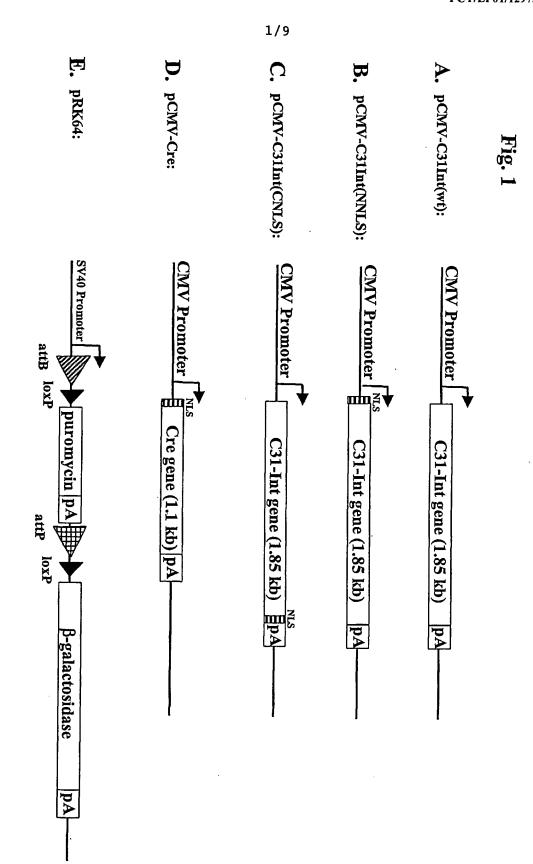
```
tcaagcattt caccttgaag ttgagacgtt ttgttagtgt atactagttt atatgttgga 11100
      ggacatgttt atccagaaga tattcaggac tatttttgac tgggctaagg aattgattct 11160
      gattagcact gttagtgagc attgagtggc ctttaggctt gaattggagt cacttgtata 11220
      tctcaaataa tgctggcctt ttttaaaaag cccttgttct ttatcaccct gttttctaca 11280
     taatttttgt tcaaagaaat acttgtttgg atctcctttt gacaacaata gcatgttttc 11340 aagccatatt tttttcctt tttttttt tttttggttt ttcgagacag ggtttctctg 11400
      tatagecetg getgteetgg aacteaettt gtagaceagg etggeetega aeteagaaat 11460
      cogectgeet etgecteetg agtgeeggga ttaaaggegt geaccaceae geetggetaa 11520
      gttggatatt ttgttatata actataacca atactaactc cactgggtgg atttttaatt 11580
     cagtcagtag tettaagtgg tetttattgg ceetteatta aaatetactg tteaetetaa 11640
      cagaggetgt tggtactagt ggcacttaag caactteeta eggatataet ageagattaa 11700
     gggtcaggga tagaaactag tctagcgtti tgtataccta ccagctttat actaccttgt 11760
      tctgatagaa atatttcagg acat
15
      <210> 107
      <211> 1458
      <212> DNA
      <213> Bacteriophage TP901-1
20
      <220>
     <221> CDS
     <222> (1)..(1455)
25
     <400> 107
     atg act aag aaa gta gca atc tat aca cga gta tcc act act aac caa
     Met Thr Lys Lys Val Ala Ile Tyr Thr Arg Val Ser Thr Thr Asn Gln
30
     gca gag gaa ggg ttc tca att gat gag caa att gac cgt tta aca aaa
                                                                          96
     Ala Glu Glu Gly Phe Ser Ile Asp Glu Gln Ile Asp Arg Leu Thr Lys
     tat gct gaa gca atg ggg tgg caa gta tct gat act tat act gat gct
35
     Tyr Ala Glu Ala Met Gly Trp Gln Val Ser Asp Thr Tyr Thr Asp Ala
     ggt ttt tca ggg gcc aaa ctt gaa cgc cca gca atg caa aga tta atc
                                                                          192
     Gly Phe Ser Gly Ala Lys Leu Glu Arg Pro Ala Met Gln Arg Leu Ile
40
                               55
     aac gat atc gag aat aaa gct ttt gat aca gtt ctt gta tat aag cta
                                                                          240
     Asn Asp Ile Glu Asn Lys Ala Phe Asp Thr Val Leu Val Tyr Lys Leu
45
     gac cgc ctt tca cgt agt gta aga gat act ctt tat ctt gtt aag gat
                                                                          288
     Asp Arg Leu Ser Arg Ser Val Arg Asp Thr Leu Tyr Leu Val Lys Asp
     gtg ttc aca aaa aat aaa ata gac ttt atc tcg ctt aat gaa agt att
                                                                          336
     Val Phe Thr Lys Asn Lys Ile Asp Phe Ile Ser Leu Asn Glu Ser Ile
                 100
     gat act tot tot got atg ggt ago ttg ttt oto act att ott tot goa
                                                                          384
55
     Asp Thr Ser Ser Ala Met Gly Ser Leu Phe Leu Thr Ile Leu Ser Ala
                                  120
     att aat gag ttt gaa aga gag aat ata aaa gaa cgc atg act atg ggt
                                                                          432
     Ile Asn Glu Phe Glu Arg Glu Asn Ile Lys Glu Arg Met Thr Met Gly
60
     aaa cta ggg cga gcg aaa tct ggt aag tct atg atg tgg act aag aca
                                                                          480
     Lys Leu Gly Arg Ala Lys Ser Gly Lys Ser Met Met Trp Thr Lys Thr
                          150
                                               155
65
     get tit ggg tat tac cac aac aga aag aca ggt ata tta gaa att gtt
                                                                          528
     Ala Phe Gly Tyr Tyr His Asn Arg Lys Thr Gly Ile Leu Glu Ile Val
```

165

| cct tta caa gct aca ata gtt ga Pro Leu Gln Ala Thr Ile Val Gl 180 gga ata tca ctt aca aaa tta ag Gly Ile Ser Leu Thr Lys Leu Ar 195 atc ggt aaa gat ata ccg tgg tc Ile Gly Lys Asp Ile Pro Trp Se 210 15 gat aat cca gtt tac tgt ggt ta Asp Asn Pro Val Tyr Cys Gly Ty 225 230 | lu Gln Ile Phe Thr Asp Ty: 185 ga gat aaa ctc aat gaa tci rg Asp Lys Leu Asn Glu Se: 00 ct tat cgt acc cta aga cas er Tyr Arg Thr Leu Arg Gli 220 at atc aaa ttt aag gac aga yr Ile Lys Phe Lys Asp Se: | E Leu Ser E gga cac 624 E Gly His A aca ctt 672 A Thr Leu |
|--|--|---|
| atc ggt aaa gat ata ccg tgg tc Ile Gly Lys Asp Ile Pro Trp Se 210 215 gat aat cca gtt tac tgt ggt ta Asp Asn Pro Val Tyr Cys Gly Ty 225 230 | rg Asp Lys Leu Asn Glu Se: 00 205 ct tat cgt acc cta aga ca: er Tyr Arg Thr Leu Arg Gli 220 at atc aaa ttt aag gac agg yr Ile Lys Phe Lys Asp Se: | Gly His a aca ctt 672 Thr Leu |
| 11e Gly Lys Asp Ile Pro Trp Se 210 215 15 gat aat cca gtt tac tgt ggt ta Asp Asn Pro Val Tyr Cys Gly Ty 225 230 | er Tyr Arg Thr Leu Arg Gli 220 at atc aaa ttt aag gac ago yr Ile Lys Phe Lys Asp Sei | Thr Leu |
| Asp Asn Pro Val Tyr Cys Gly Ty 225 230 | yr Ile Lys Phe Lys Asp Sei | cta ttt 720 |
| | 235 | Leu Phe 240 |
| 20 Glu Gly Met His Lys Pro Ile Il 245 | tc cct tat gag act tat ttd le Pro Tyr Glu Thr Tyr Let 250 | a aaa gtt 768 1 Lys Val 255 |
| caa aaa gag cta gaa gaa aga ca Gln Lys Glu Leu Glu Glu Arg Gl 25 260 | aa cag cag act tat gaa aga ln Gln Gln Thr Tyr Glu Arc 265 270 | J Asn Asn |
| aac cct aga cct ttc caa gct aa Asn Pro Arg Pro Phe Gln Ala Ly 275 | ys Tyr Met Leu Ser Gly Met | g gca agg 864 : Ala Arg |
| tgc ggt tac tgt gga gca cct tt Cys Gly Tyr Cys Gly Ala Pro Le 290 295 | ta aaa att gtt ctt ggc cad eu Lys Ile Val Leu Gly His 300 | c aaa aga 912 s Lys Arg |
| 35 aaa gat gga agc cgc act atg aa Lys Asp Gly Ser Arg Thr Met Ly 305 310 | aa tat cac tgt gca aat aga ys Tyr His Cys Ala Asn Arc 315 | ttt cct 960 g Phe Pro 320 |
| cga aaa aca aaa gga att aca gt Arg Lys Thr Lys Gly Ile Thr Va 325 | ta tat aat gac aat aaa aaq al Tyr Asn Asp Asn Lys Lys 330 | tgt gat 1008 Cys Asp 335 |
| tca gga act tat gat tta agt aa Ser Gly Thr Tyr Asp Leu Ser As 340 | at tta gaa aat act gtt att sn Leu Glu Asn Thr Val Ile 345 350 | Asp Asn |
| ctg att gga ttt caa gaa aat aa Leu Ile Gly Phe Gln Glu Asn Ass 355 360 | sn Asp Ser Leu Leu Lys Ile | atc aat 1104 :Ile Asn |
| ggc aac aac caa cct att ctt ga Gly Asn Asn Gln Pro Ile Leu Asp 370 | at act tcg tca ttt aaa aag sp Thr Ser Ser Phe Lys Lys 380 | caa att 1152 Gln Ile |
| 55 tca cag atc gat aaa aaa ata ca Ser Gln Ile Asp Lys Lys Ile Gl 385 390 | na aag aac tct gat ttg tac .n Lys Asn Ser Asp Leu Tyr 395 | cta aat 1200 Leu Asn 400 |
| gat ttt atc act atg gat gag ttc Asp Phe Ile Thr Met Asp Glu Let 405 | g aaa gat cgt act gat tcc u Lys Asp Arg Thr Asp Ser 410 | ctt cag 1248 Leu Gln 415 |
| gct gag aaa aag ctg ctt aaa gc Ala Glu Lys Lys Leu Leu Lys Ala 65 420 | et aag att agc gaa aat aaa .a Lys Ile Ser Glu Asn Lys 425 430 | Phe Asn |
| gac tot act gat gtt ttt gag tta | a gtt aaa act cag ttg ggc | tca att 1344 |

| | ` | 0.00 | ./ JUU 1 | | | | | | | 00 | | | | | | PC1/ | EP01/129 |
|----|-------------------|----------------------------------|------------|------------|------------|-------------------|-------------------|------------|------------|------------|-------------------|-------------------|------------|------------|------------|-------------------|----------|
| | Asp | Ser | Thr 435 | Asp | Val | Phe | Glu | Leu 440 | Val | 93 Lys | Thr | Gln | Leu 445 | Gly | Ser | Ile | |
| 5 | ccg Pro | att Ile 450 | aat Asn | gaa Glu | cta Leu | tca Ser | tat Tyr 455 | gat Asp | aat Asn | aaa Lys | aag Lys | aaa Lys 460 | atc Ile | gtc Val | aac Asn | aac Asn | 1392 |
| 10 | ctt Leu 465 | gta Val | tca Ser | aag Lys | gtt Val | gat Asp 470 | gtt Val | act Thr | gct Ala | gat Asp | aat Asn 475 | gta Val | gat Asp | atc Ile | ata Ile | ttt Phe 480 | 1440 |
| 15 | | | caa Gln | | | taa | | | | | | | | | | | 1458 |
| 20 | <213 <213 | 0> 10 1> 48 2> PI 3> Ba | 35 | riopl | nage | TP9(|)1-1 | | | | | | | | | | |
| 25 | | 0> 1(Thr | 08 Lys | Lys | Val 5 | Ala | Ile | Tyr | Thr | Arg 10 | Val | Ser | Thr | Thr | Asn 15 | Gln | |
| 23 | Ala | Glu | Glu | Gly 20 | Phe | Ser | Ile | Asp | Glu 25 | Gln | Ile | Asp | Arg | Leu 30 | Thr | Lys | |
| 30 | Tyr | Ala | Glu 35 | Ala | Met | Gly | Trp | Gln 40 | Val | Ser | Asp | Thr | Tyr 45 | Thr | Asp | Ala | |
| | Gly | Phe 50 | Ser | Gly | Ala | Lys | Leu 55 | Glu | Arg | Pro | Ala | Met 60 | Gln | Arg | Leu | Ile | |
| 35 | Asn 65 | Asp | Ile | Glu | Asn | Lys 70 | Ala | Phe | Aśp | Thr | Val 75 | Leu | Val | Tyr | Lys | Leu 80 | |
| 40 | Asp | Arg | Leu | Ser | Arg 85 | Ser | Val | Arg | Asp | Thr 90 | Leu | Tyr | Leu | Val | Lys 95 | Asp | |
| | Val | Phe | Thr | Lys 100 | Asn | Lys | Ile | Asp | Phe 105 | Ile | Ser | Leu | Asn | Glu 110 | Ser | Ile | |
| 45 | | | Ser 115 | | | | | 120 | | | | | 125 | | | | |
| 50 | | 130 | Glu | | | | 135 | | | | | 140 | | | | _ | |
| 50 | 145 | | Gly | | | 150 | | | | | 155 | | | | | 160 | |
| 55 | | | Gly | | 165 | | | | | 170 | - | | | | 175 | | . • |
| | | | Gln | 180 | | | | | 185 | | | | | 190 | | | |
| 60 | | | Ser 195 | | | | | 200 | | | | | 205 | | | | |
| 65 | | 210 | Lys | | | | 215 | | | | | 220 | | | | | |
| 65 | Asp 225 | Asn | Pro | Val | Tyr | Cys 230 | Gly | Tyr | Ile | Lys | Phe 235 | Lys | Asp | Ser | Leu | Phe 240 | |

Lys Phe Gln Leu Ala



() ()

relative activity 0.004 0.23 0.32 0.36 0.3 0.65 0.5 0.63 0.72 5 x 104 RLU x105 (Gal/Luc) -\$-RLU x106 (Gal/Luc) 46975 ± 3696 23299 ± 3194 33872 ± 2609 14412 ± 2050 2677621 ± 504285 | 17029 ± 2246 30560 ± 1585 29595 ± 4632 2560450 ± 736186 | 15104 ± 3041 10774 ± 972 204 ± 62 24 ± 4 1657062 ± 526562 1173932 ± 291315 868905 ± 196404 RLU (Luciferase) 1818700 ± 328970 763121 ± 280687 677861 ± 145341 988144 ± 175116 864752 ± 229935 527237 ± 53846 784932 ± 290524 274192 ± 78937 231200 ± 96741 262169 ± 60583 6) pCMV-C31Int(NNLS) 0.5 ng 158402 ± 75870 206857 ± 76733 297760 ± 83363 75838 ± 12628 73318 ± 19084 RLU (BGal) 1936 ± 262 433 ± 37 8) pCMV-C31Int(CNLS) 0.5 ng 0.5 ng 0.5 ng 7) pCMV-C31Int(NNLS) 1 ng 9) pCMV-C31int(CNLS) 1 ng 1 ng 5) pCMV-C31Int(wt) 1 ng 4) pCMV-C31Int(wt) 10) pCMV-Cre(NNLS) 11) pCMV-Cre(NNLS) 2) pRK64(\u00e4Cre) 1) pUC19 only 3) pRK64 only Sample Fig.

SUBSTITUTE SHEET (RULE 26)

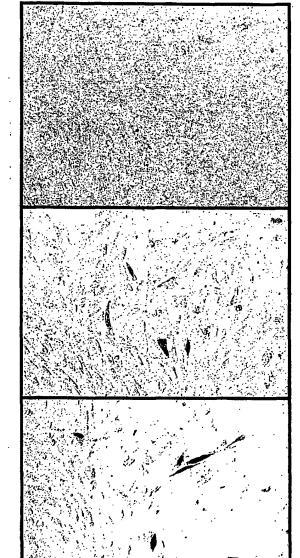
F16. 3

| | RLU (βGal) | RLU (Luciferase) | RLU x 10 ⁵ (Gal/Luci) |
|-----------------------------|--------------|-------------------|----------------------------------|
| 1) pPGKnifD (reporter) only | 1324 ± 876 | 3631598 ± 903012 | 34 ± 18 |
| 2) pCMV-XisA 25 ng | 4650 ± 2273 | 2741969 ± 667568 | 164 ± 54 |
| 3) pCMV-XisA 100 ng | 17529 ± 9304 | 3798872 ± 1288020 | 443±151 |
| 4) pCMV-XisA(NNLS) 25 ng | 4060 ± 1376 | 2471695 ± 611351 | 163 ± 36 |
| 5) pCMV-XisA(NNLS) 100 ng | 17801 ± 3892 | 3570103 ± 750628 | 500 ± 65 |
| 6) pPGKattA (reporter) only | 754 ± 70 | 195822 ± 81858 | 755±601 |
| 7) pCMV-SSV 10 ng | 925±273 | 119043 ± 67451 | 906 ± 316 |
| 8) pCMV-SSV 20 ng | 1033±270 | 122557 ± 30054 | 879 ± 291 |
| 9) pCMV-SSV(NNLS) 10 ng | 1108 ± 367 | 174380 ± 58876 | 694.± 345 |
| 10) pCMV-SSV(NNLS) 20 ng | 1306 ± 383 | 211182 ± 101011 | 874 ± 741 |
| | | | |

9) pCMV-Cre(NNLS) 8) pCMV-Cre(NNLS) 7) pCMV-C31Int(CNLS) 64 ng 6) pCMV-C31Int(CNLS) 32 ng 5) pCMV-C31int(NNLS) 64 ng 4) pCMV-C31Int(NNLS) 32 ng 3) pCMV-C31Int(wt) 2) pCMV-C31Int(wt) 1) pUC19 only Sample 64 ng 64 ng 32 ng 32 ng 36823 ± 3993 27064 ± 3769 28849 ± 6623 17624 ± 5578 8206 ± 2210 3783 ± 1537 7125 ± 1474 1386 ± 174 754 ± 44 RLU (βGal) 5 x 103 RLU (βGal) 4 x 10⁴ relative activity 0.73 0.48 0.22 0.78 0.19 0.1 0.04

Fig. 4

A.: Nontransfected control

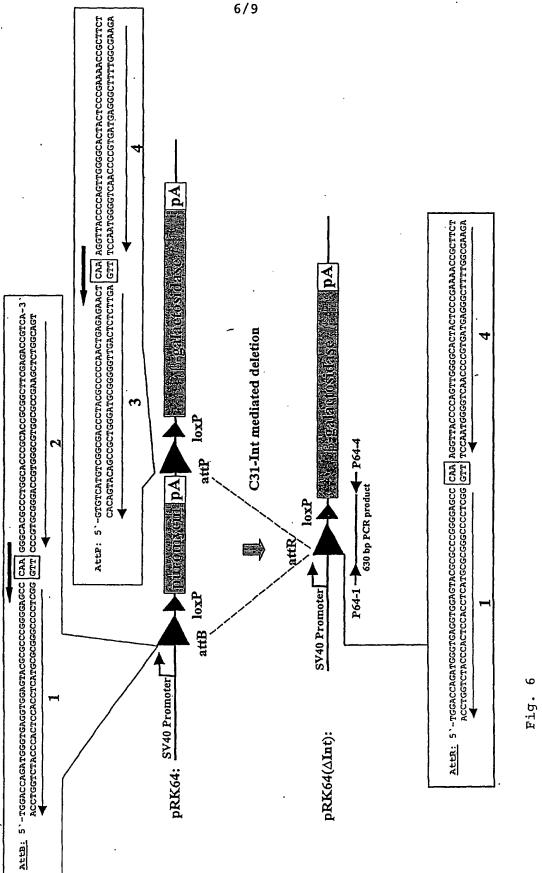


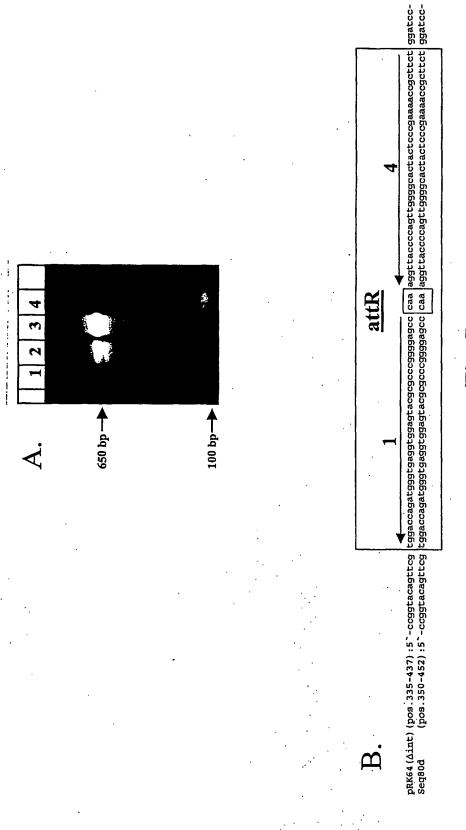
 $B.: \, \mathsf{pCMV\text{-}Cre}$

C.: pcmv-c31Int(NLS)

Fig. 5

4 C +





SUBSTITUTE SHEET (RULE 26)

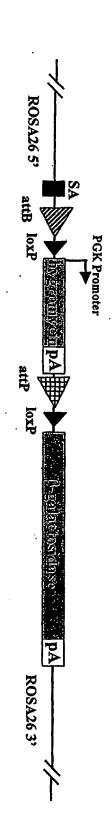


Fig. 8

9/9

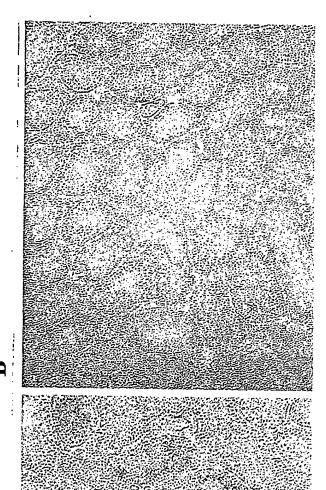


Fig. 9